DOCUMENT RESUME

ED 436 623 CE 079 232

TITLE	Cross-Industry Assessment and Certification. Framework and Implementation Guide.
INSTITUTION	WestEd, San Francisco, CA.
SPONS AGENCY	Department of Labor, Washington, DC.; Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	1998-01-00
NOTE	136p.; Also supported by funding from the National Skills Standards Board.
CONTRACT	RJ96006901
PUB TYPE	Guides - Non-Classroom (055) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Adult Education; Education Work Relationship; Educational Certificates; Elementary Secondary Education; Higher Education; *Job Skills; Models; *Performance Based Assessment; *Portfolio Assessment; Postsecondary Education; School Business Relationship; Scoring; *Standards; *Student Certification; *Student Evaluation; Vocational Education
IDENTIFIERS	National Skill Standards Board

ABSTRACT

This guide provides assessment models to those involved in planning and implementing workplace assessment and certification systems, including systems supported by the School-to-Work Opportunities Act. Chapter 2 presents the following: a kindergarten-through-career vision for cross-industry assessment; a plan for spanning the needs and goals of industry and education contexts; an assessment of strategies for each developmental level; and ways in which this vision is aligned to and supports the National Skill Standards Board (NSSB) vision for industry-skill standards. Chapter 3 defines standards, presents types of standards used in education and industry contexts, and provides a summary of work done by the National Skill Standards and Assessment Collaborative (NSSAC) to identify categories of cross-industry standards applicable across industries. Chapter 4 describes qualities of an effective program and discusses validity, reliability, and fairness. Chapter 5 charts NSSAC's development work with two assessment methods--portfolios and written scenarios. The identified assessment prototypes are included. Chapter 6 presents a general process for developing effective scoring systems, highlighted by specific examples, and discusses issues related to combining assessment information and reporting results to appropriate audiences. Chapter 7 summarizes "lessons learned" and recommendations for design and implementation of a skill standards-based assessment system. Appendixes include information on NSSB and NSSAC scenario review form. The guide contains 52 references. (YLB)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 436 623

national skill standards & assessment collaborative

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CE079232

Cross-Industry Assessment and Certification

Framework and Implementation Guide

BEST COPY AVAILABLE

Cross–Industry Assessment and Certification

Framework and Implementation Guide

January 1998

National Skill Standards and Assessment Collaborative

About WestEd

WestEd is a non-profit research, development and service agency dedicated to improving education and other opportunities for children, youth and adults. Drawing on the best from research and practice, we work with practitioners, policymakers and others in the broader education community to address critical issues in education and other related areas: from early childhood intervention and support to school-to-work transition, from standards and assessment development to safe schools and safe communities. In addition to its research and development work, WestEd offers program evaluation, policy analysis, professional development and other key services.

The agency was created in 1995 to unite and enhance the capacity of Far West Laboratory and Southwest Regional Laboratory, two of the nation's original education laboratories created by Congress in 1966. In addition to our work across the nation, WestEd serves as the regional education laboratory for Arizona, California, Nevada and Utah. The organization is headquartered in San Francisco, with additional offices in Arizona, Massachusetts, Washington, D.C. and elsewhere in California.

About Assessment and Standards Development Services

From the local to the national level, WestEd's Assessment and Standards Development Services (ASDS) is involved extensively in planning, developing, implementing and evaluating tools and systems for ensuring quality and accountability, in both education and the workplace. ASDS addresses standards and assessment issues within such traditional academic areas as language arts, mathematics, science and social science, as well as in such emerging fields as school-to-work transition and teacher preparation and credentialing. In its assessment work, ASDS staff seek ways to combine traditional and innovative methods into coherent, affordable and valid assessment systems that provide the feedback needed for accountability purposes or to help guide teaching and learning in the classroom or the workplace.

The development and production of this document was supported by funding from the National Skills Standards Board (NSSB) and the U.S. Department of Labor. Additional funding came from the U.S. Department of Education, Office of Educational Research and Improvement, contract number RJ96006901. Its contents do not necessarily reflect the views or policies of the NSSB or the U.S. Department of Education, nor does mention of trade names, commercial products or organizations imply endorsement by any of the above.

Assessment and Standards
Development Services
WestEd
730 Harrison Street
San Francisco, CA 94107-1242
Telephone: (415) 241-2725
Fax (415) 241-2702
www.WestEd.org

© 1998 WestEd. All rights reserved.

Table of Contents

Chapter 1: Introduction ____ 1

Purpose and Goals of the *National Skill Standards and Assessment Collaborative (NSSAC) Cross-Industry Assessment and Certification Guide* ____ 1

The National Skill Standards and Assessment Collaborative (NSSAC) ____ 1

Relevant Legislation Supporting NSSAC's Work ____ 2

How the Guide is Organized ____ 3

Chapter 2: A Comprehensive Vision for Cross-Industry Assessment and Certification ____ 5

An Ideal Skills Assessment System ____ 5

Elementary School ____ 6

Middle School ____ 7

High School ____ 7

Grade 10 ____ 7

Grade 12 ____ 8

Higher Education ____ 10

Workplace Skills Assessment ____ 11

Benefits of a Comprehensive Vision ____ 12

Chapter 3: Standards-Based Reform for Education and Industry ____ 15

Education and Industry as Partners ____ 15

Defining Standards ____ 16

Academic Standards ____ 19

Employability Skill Standards ____ 20

Industry Skill Standards ____ 21

Standards Development / Adaptation and Validation Processes ____ 23

NSSAC's Efforts ____ 25

NSSAC's Cross-Industry Analysis of Standards ____ 25

NSSAC Cross-Industry Categories of Standards ____ 29

Chapter 4: Qualities of an Effective Occupational Assessment and Certification Program ____ 33
Assessment and Certification for the Workplace ____ 33
Ensuring the Technical Quality of an Assessment and Certification Program ____ 33
Is the Assessment Aligned to Challenging Standards? ____ 34
Does the Assessment Incorporate Multiple Measures? ____ 35
Is the Assessment Valid and Does it Reflect What is Valued in the Workplace? ____ 40
Is the Assessment Reliable and Free from Measurement Error? ____ 42
Are Scores Comparable Across Different Sites? ____ 44
Is the Assessment Equitable and Fair? ____ 45
Is the Assessment Practical and Cost Effective? ____ 47
Establishing an Assessment and Certification Program ____ 47

Chapter 5: NSSAC's Cross-Industry Assessment Prototypes ____ 49
The Written Scenario ____ 50
The Process Used to Develop the Written Scenario Prototypes ____ 51
NSSAC's Pilot-Testing Efforts ____ 55
Issues in Developing Cross-Industry Written Scenarios ____ 57
The Written Scenario Prototypes ____ 59
Portfolios ____ 68
Educational Purposes for Portfolios ____ 69
Industry Purposes for Portfolios ____ 76
A Cross-Industry Portfolio Template ____ 79
Developing the Cross-Industry Portfolio Template ____ 79
Expert Panel Groups ____ 79
Employer Survey ____ 80
The Cross-Industry Portfolio Template ____ 82

Chapter 6: Scoring and Reporting ____ 84
Developing a Scoring System ____ 84
Scoring Written Scenario Assessments — the NSSAC Example ____ 95
Reporting Assessment Results ____ 104

Chapter 7: Lessons Learned and Implications for Designing a Comprehensive Skill Standards-Based Assessment System ____ 106
Lessons Learned ____ 106
Conditions for Successful Implementation of Skill Standards-Based Assessment Systems ____ 109
Concluding Remarks ____ 111

Appendix A: The National Skill Standards Board (NSSB) ____ 113
The National Skill Standards Board (NSSB) ____ 113
The NSSB Framework for Industry Skill Standards ____ 113

Appendix B: NSSAC Scenario Review Form ____ 117

Appendix C: Employer Survey Results ____ 119

NSSAC Project Partners ____ 121

NSSAC Pilot Sites ____ 123

Bibliography ____ 125

NSSAC Project Staff ____ 130

Chapter 1: Introduction

"From now on the U.S. will only be competitive if we all work smarter and we employ people with more highly developed skills."

— James R. Houghton
Chairman, National Skill Standards Board
Retired Chairman and CEO, Corning Incorporated

Purpose and Goals of the *National Skill Standards and Assessment Collaborative (NSSAC) Cross-Industry Assessment and Certification Guide*

The *NSSAC Cross-Industry Assessment and Certification Framework and Implementation Guide* is intended as a resource for national and state policymakers, employers, educators, and assessment developers who are interested in standards-based, career-related assessment systems. In particular, the guide is designed to provide viable assessment models to those involved in planning and implementing workplace assessment and certification systems, including those systems supported by the School-to-Work Opportunities Act of 1994 (STWOA).

The *Guide* provides (1) a conceptual framework for cross-industry assessment (i.e., assessment methodologies that apply across industries), (2) specific examples of assessments for use in education, training, and the workplace, and (3) guidelines for developing assessment and certification systems. Informed by the National Skill Standards Board (NSSB), the conceptual framework for NSSAC cross-industry assessments is designed to support system-building efforts at the local, state, and national levels.

The National Skill Standards and Assessment Collaborative (NSSAC)

The National Skill Standards and Assessment Collaborative (NSSAC) combines the efforts of four national skill standards pilot projects to address issues concerning assessment and certification. With WestEd as the lead partner, NSSAC includes four diverse industries — health care (represented by WestEd), electronics (American Electronics Association, AEA), human services (Human Services Research Institute, HSRI), and retail (National Retail Federation, NRF). With support from the U. S. Department of Labor (DOL) and the National Skill Standards Board (NSSB), NSSAC has (1) identified cross-industry standard areas or categories that are linked to the industry skill standards developed for the four project partners, and (2) pilot tested assessment methods that apply across these industries, based on these standards.

The NSSB framework (see **Appendix A** for a detailed discussion of this framework) is the basis from which NSSAC began its work on cross-industry assessment and certification. By providing a

framework that includes both broad and specialized levels, the NSSB has prepared a strong foundation for design of a cross-industry, as well as occupation-specific, assessment and certification system. Ideally, such a system can be used to align education and industry interests by providing cross-industry categories that are essential for industry, applicable to classroom instruction, and linked to a high level of challenge and expectation.

Relevant Legislation Supporting NSSAC's Work

The research and development conducted by NSSAC grows out of a wave of legislation aimed at aligning education and industry interests to better meet the needs of students, workers, and employers. The *Carl D. Perkins Vocational and Applied Technology Education Act (Perkins II)* of 1990 was pivotal in initiating the process of broad-based consensus building between education and industry. This law required vocational education programs to develop and implement a system of performance standards, assessment measures, and services that provide "strong experience in and understanding of all aspects of the industry students are preparing to enter...." (Perkins II).

Following Perkins II in 1992, the U. S. Departments of Education and Labor jointly initiated funding for projects to develop industry skill standards in 22 diverse industries. The NSSAC partners represented four of these original 22 projects. Several interlocking pieces of legislation in subsequent years further strengthened the call for integrated education and industry standards: the *School-to-Work Opportunities Act (STWOA), Goals 2000: Educate America Act, Improving America's Schools Act, and National Skill Standards Act* (see **Appendix A** for a discussion of the National Skill Standards Board which was established by this legislation). Taken together, these legislative mandates promote the development of voluntary systems of national academic and industry skill standards and assessments.

Presently, the NSSB is funding "voluntary partnerships" with broad member representation to endorse skill standards in different economic sectors. Now that the 22 original national skill standards projects have demonstrated viable processes for developing and validating national industry-specific standards, next steps include exploring commonalities across industries. Working toward commonalities will inform development of standards and assessment systems for the broader economic sectors.

The purpose of this WestEd led project is to: (a) identify common cross-industry standard areas derived from our project partners' industry skill standards, and (b) subsequently pilot test assessment methods that target these common areas as well as individual skill standards. This guide is a culmination of that work.

How the Guide is Organized

Subsequent parts of this guide are organized as follows:

Chapter 2: A Comprehensive Vision for Cross-Industry Assessment and Certification
This chapter presents a Kindergarten-through-Career vision for cross-industry assessment. The chapter proposes a plan for spanning the needs and goals of industry and education contexts and offers specific assessment strategies for each developmental level, beginning in elementary school and culminating in the workplace. A final section discusses how this vision is aligned to and supports the NSSB vision for industry skill standards.

Chapter 3: Standards-Based Reform for Education and Industry
This chapter addresses standards, what they are, and their pivotal role in reform efforts; particularly in laying the groundwork for a comprehensive assessment system. It presents the different kinds of standards used in education and industry contexts, with an emphasis on industry skill standards. The chapter concludes with a discussion of NSSAC's analysis of the project partners' industry skill standards and the eight cross-industry standards categories that form the basis of preliminary cross-industry assessment development.

Chapter 4: Qualities of an Effective Assessment and Certification Program
The purpose of the chapter is to provide readers with background information about the technical qualities of assessment and certification programs, including discussions of validity, reliability, and fairness.

Chapter 5: NSSAC's Cross-Industry Assessment Prototypes
This chapter charts NSSAC's development work with two assessment methods — portfolios and written scenarios. The purpose of the development effort was to identify assessment prototypes, which are included in the chapter. These prototypes can be

applied across many industry contexts and form the basis for cross-industry assessment.

Chapter 6: Scoring and Reporting
This chapter presents a general process for developing effective scoring systems highlighted by specific examples. Issues related to combining assessment information and reporting results to appropriate audiences are also discussed.

Chapter 7: Lessons Learned and Implications for Designing a Comprehensive Skill Standards-Based Assessment System
This concluding chapter summarizes "lessons learned" and specific recommendations for design and implementation of a skill standards-based assessment system.

Chapter 2: A Comprehensive Vision for Cross-Industry Assessment and Certification

"Skill standards will help us develop the next generation of retail workers — a committed, professional workforce that takes pride in what it can do."

— Katherine T. Mance
Vice President, Research, Education, and Community Affairs
National Retail Federation

An Ideal Skills Assessment System

What is needed to support the NSSB framework for industry skill standards is a comprehensive vision for assessment and certification that spans elementary through postsecondary education, culminates at the workplace, and applies across various industries. The work of the National Skill Standards and Assessment Collaborative (NSSAC) was intended to help shape and inform that vision.

An ideal skills assessment system involves assessing performance at key junctures in an individual's education and training in order to reinforce and measure that individual's proficiency with respect to agreed upon standards. In the educational arena, assessment development and implementation should always involve a partnership of educators and industry representatives (as well as parents and the broader community). However, starting in high school, industry must take on more responsibility and ultimately assume the lead. This is consistent with the school-to-work and skill standards movements, both of which espouse that standards and assessments must be industry-driven in order to have credibility and validity for career applications.

Described in detail elsewhere (Ananda & Rabinowitz, 1995; Wills, 1997; Ananda [in progress]), this model is graphically presented below in **Table 1**.

Table 1.
An Ideal Skills Assessment System

	Targeted Skills						
	Academic	General Workplace	Industry Foundation or Core	Occupational Family	Occupational Specific	Certification Status	Assessment Partners*
Elementary School	X	X				None	*Education* / Industry
Middle School	X	X				None	*Education*/ Industry
Grade 10	X	X	X			Certificate of Initial Mastery	*Education* / Industry
Grade 12	X	X	X	X		Certificate of Advanced Mastery	*Education*/ *Industry*
Higher Education	X	X	X	X	X	Career Specialization	*Industry* / Education
Workplace	X	X	X	X	X	Job Entry / Reentry Career Specialization	*Industry*/ Education

* lead partner for assessment development and implementation is indicated by italics

A description of the above framework for each developmental level follows.

Elementary School

Targeted skills. Students should begin to understand the importance of developing employability skills and positive work attitudes. At this point, instruction should introduce students to the notion of there being a variety of work and careers, the importance of quality in the workplace, and some understanding of what it takes to be successful in the workplace. For example, students should be able to use skills from more than one discipline to complete a project organized around a work or career-related theme. These types of activities will help students to develop an appreciation for the importance and dignity of work and workers.

Assessment Certification Process. At this level, assessment of targeted skills can occur as part of the ongoing testing program and assessment of workplace readiness skills in the context of specific curriculum-related tasks or projects. Assessment should not entail certification or sanctions for students.

Assessment Partners. Although industry may have an active role in schooling through involvement in learning activities, industry has minimal involvement in student assessment. Educators have primary responsibility with support from the family for development of academic and workplace readiness skills.

Middle School

Targeted Skills. With career awareness instruction beginning in elementary school, assessment of career-related skills should start by middle school. At this point, instruction should be on building academic foundations required for full participation in society, including employment and citizenship. The academic curriculum should also infuse career-related themes and concepts, such as quality control in science education. To reinforce such learning, assessment should focus on these academic and life foundation skills that are often excluded from the traditional academic curriculum, such as teamwork and personal responsibility.

Assessment and Certfication Process. As in the earlier grades, assessment of the targeted skills can occur as part of the ongoing academic testing program and should not entail formal certifications or sanctions for students. In addition to traditional paper-and-pencil assessment tasks, use of assessments such as portfolios and teamwork exercises at this level should help reinforce both academic and workplace readiness skills.

Assessment Partners. Primary responsibility for assessment remains at the local rather than the state level, with educators in the lead. Although formal career-related certification is considerably premature at this level, community or industry groups should participate along with educators in development or review of the assessments. This will help facilitate the infusion of real-life applications with academic content.

High School

Two distinct levels of formal assessment and certification should be offered in high school, as described below.

Grade 10

Targeted Skills. Many states stipulate that students should complete their core training in key academic subjects (e.g., mathematics, science, language arts). As with the middle school level,

assessments at grade 10 should focus more heavily on academic concepts than industry or career skills. Nevertheless, students at this level should be exposed to generic workplace readiness or employability skills, such as those identified by the SCANS (see **Figure 4** on page 21). Such skills are often introduced in career-technical foundation courses, such as business administration core or health careers core. Assessments at this level should reinforce attainment of these generic employability and industry core skills.

Assessment and Certfication Process. Some states are planning to issue Certificates of Initial Mastery (CIM) to recognize achievement in key academic subjects, thereby shifting substantial responsibility for development of formal assessments from the local to the state level. Comprehensive school-to-work systems should expand upon current CIM assessment models from their core academic base to include assessment of generic workplace readiness or employability skills and the introduction of career preparation focusing on a particular economic sector or industry. The primary objective of the expanded CIM is to help ensure that students are exposed to information about broad career options early on in high school and hence do not limit their future options due to inadequate information or academic training. Again, a combination of traditional paper-and-pencil tasks as well as more innovative assessment techniques would allow the student to demonstrate mastery in different ways.

Assessment Partners. Educators should take the lead in assessment development and implementation, although assessment development teams should also include parent, civic, and industry representation in order to secure valuable input as well as the necessary "buy in" for the certification process.

Grade 12 (Exit)

Targeted Skills. The focus should be on advanced academic skills in the context of a specific career area (e.g., electronics manufacturing, agriculture science). Many recent school-to-work models stipulate that academic training after grade 10 should be integrated within the context of a broad occupational area (e.g., health science, production technology, business and management) or career cluster. For example, career academies (Stern, Raby, and Dayton, 1992; Linnehan, 1996), built around industries ranging from health to electronics, represent a specialized attempt to combine academic and career training into meaningful student-based experiential learning.

In general, high school exit assessments should focus on broad career preparation within or across industries, including strong emphases on self-directed learning and academic achievement. This broader preparation will better serve youth in the long run because it will give them the flexibility needed to adjust to the changing workplace and accompanying changes in specific job demands and requirements. For a high-school student, job-specific training may be too narrow and may limit the student's later options. With a few notable exceptions, occupation-specific assessment and training should be the focus of higher education and the workplace, rather than high school.

Assessment and Certfication Process. Whereas grade 10 certification focuses on academic foundations, generic workplace readiness, and industry core skills, at or around grade 12 students could pursue a more advanced level of certification that demonstrates mastery of a more specialized level of academic and technical knowledge and skill relevant to a cluster of related occupations, or an occupational family. States such as Oregon refer to this level of high school certification as Certificates of Advanced Mastery (CAM). Whereas Oregon's CAM currently focuses more on advanced academic rather than career specialization skills, the concept of the expanded CAM presented here would actively support both options. That is, CAMs could be offered in humanities (e.g., for college-bound liberal arts majors) or occupational families, such as graphic design, bioscience, or electronics. In Oregon, development of endorsement credentials are currently underway, providing an option for students to be recognized for in-depth, career-related study.

In general, the Certificate of Advanced Mastery would allow for greater use of performance-based assessment than Certificate of Initial Mastery. First, fewer students may be prepared to pursue advanced certification. Lower numbers can mean more manageable assessment administration and scoring demands, thereby making performance-based assessment a more feasible option. Moreover, the more specialized preparation within an occupational family provides greater opportunities for career-related, performance-based assessment. For example, students at this level have more career-related, content knowledge to develop long-term projects (e.g., patient case studies in health careers or comprehensive marketing plans for business administration) that can serve as assessment tasks. Finally, students' advanced training often includes paid or unpaid on-the-job experience which

also lends itself to more "hands on" performance assessment (e.g., job/work samples).

Assessment Partners. As compared to earlier grade levels, the career specialization at grade 12 suggests that industry must take a more active role in assessment development and implementation. At this point, educators and industry representatives must be equal partners. In this way, assessments can be closely tied with industry expectations but still have fidelity to classroom curriculum and practices. In addition, parents and the broader community should be involved in the assessment development and implementation process.

Higher Education

Targeted Skills. A traditional focus for community college curriculum is on occupational specific training (e.g., nursing assistant, welding, drafting). However, training at this level is also broadening to include fundamental preparation in a specific industry (core curriculum) or occupational family to help prepare students for horizontal and vertical movements in their chosen careers. In addition to high level academic skills, assessments at this level should encompass general workplace readiness, occupational family, and occupational specific skills, as appropriate.

The four-year colleges and universities traditionally have played a less active role in articulated career preparation. Career preparation is often perceived as inconsistent or even at odds with a liberal arts curriculum. However, the STW movement is beginning to make important in-roads in involving such institutions in state and local STW partnerships.

Assessment and Certfication Process. Certification should involve multiple modes of assessment, for all the reasons alluded to previously.

Assessment Partners. At this level, the links between education and industry need to be strongest because this is the major point of transition into careers for an increasing number of individuals. For the first time in the educational/training process, industry should drive the assessment development and implementation process, with full participation of educators and other stake holders.

Workplace Skills Assessment

Targeted Skills. Success in the workplace requires proficiency on a full range of skills, ranging from occupational specific to related academics.

Assessment and Certfication Process. Entry into many occupations is regulated through certification and licensure processes. The assessments used for certification and licensure typically are called for and supported by professional associations, as well as the state and federal governments.

Many well-established certification programs — such as the National Institute for Automotive Service Excellence and the American Institute of Certified Public Accountants — rely heavily on multiple-choice assessment because of the many advantages associated with this assessment methodology. Multiple-choice items afford broad coverage of content over a relatively short amount of time, have a strong empirical history with demonstrated technical quality, and are easy and relatively inexpensive to administer. In contrast, many programs have found performance-based assessments to be preferable because they more closely evaluate the kinds of knowledge, skills, and abilities required in the workplace. Performance-based assessment is an assessment technique that requires respondents to go beyond selection of the correct answer to production of the correct answer. Several existing certification systems augment their multiple-choice components with essay or selected performance-based assessment tasks, drawing on the benefits of both kinds of assessment strategies.

Job entry is but one application of assessment for the workplace. For example, training or retraining is being offered at the workplace for new and incumbent workers. Assessment at the end of training helps to gauge the effectiveness of the training as well as the skill level of the individual worker. As work site training is often short term in nature (e.g., two days to si months), incorporation of time-intensive assessment tasks, such as portfolios, may not be feasible. Instead, on-demand or "on the spot" assessments that can be administered in a limited period of time may be more appropriate. This could include multiple-choice examinations, essay-type examinations, or direct observations of work.

Likewise, workers in any industry are subject to periodic review, typically by their immediate supervisors and sometimes by their

peers. Performance review should be based on explicit criteria of which both the supervisor and employee are aware, such as industry skill standards. Finally, assessment of an individual's skills can be used as evidence to help "make the case" for an individual's career advancement or promotion. For both performance review and promotion, a collection of an individual's work samples, such as a portfolio, could provide a useful framework for assessment.

Assessment Partners. For all workplace assessment applications, industry clearly needs to be the driver in developing and administering such programs. However, education should participate in certain applications, such as certification for job entry. Involving educators would allow them to better understand the skills coveted by industry. They can then bring back this knowledge to help ensure that what occurs during the formal education process is better linked to workplace needs.

Benefits of a Comprehensive Vision

A major objective of the NSSB is to develop "a flexible, coherent system of assessments, appropriate for settings as diverse as schools and industries" (Ananda and Rabinowitz, 1996). In keeping with the goals of the NSSB, the certification and assessment systems should reflect industry needs for a trained and adaptable workforce while drawing on expertise from educators, parents, and other members of the community.

The cross-industry assessment prototypes contained within this guide have been designed to support this NSSB mission and framework. These prototypes provide an essential focal point for communication among those working in education, school-to-work, and workplace training settings. That is, industry representatives working in collaboration with educators can adapt the prototypes to both workplace and classroom settings by (1) integrating industry themes with academic skills and (2) providing students and workers with the opportunity to apply knowledge to real-life work situations. These assessments then form the basis for dialogue around the standards and around creating a systemic vision for workplace preparation.

This integration of industry themes into academic curricula can occur in *any* educational setting, whether students are college-bound or more interested in gaining immediate entrance into the workforce. For all students, it is meaningful and relevant to

apply academic skills to solving realistic problems that reflect the world they will enter. Equally, for all students, it is more beneficial to focus on understanding, use, and application of knowledge and skills than solely on remembering discrete parts of knowledge or skill. As the assessment prototypes are based on challenging standards, they should be rich and complex enough to adapt to the full range of grade levels and school settings and to measure complex, integrated knowledge and skills.

For the reasons above, a comprehensive cross-industry assessment and certification program can help to close the gap between education and industry contexts in terms of their goals and how assessments are used to support these goals. Cross-industry assessment and certification programs span the needs of both contexts. They:

- support upward progress through the grade levels and into the workforce by teaching the very skills they measure;
- go beyond factual recall and link to complex, integrated skills such as problem-solving, critical thinking, communication and use of technology; and
- employ a variety of assessment strategies (such as scenarios and portfolios) to assess relevant knowledge, skills, and abilities.

Such a system envisions students as prospective workers and workers as ongoing students. Within this system, assessments can both measure progress, supporting the developmental nature of education and training, and provide an accurate picture of workforce readiness, supporting industry needs for immediate competence and an efficient training protocol (Ananda, in progress).

NSSB also demands an assessment protocol that supports the need for *accountability* of workplace training programs. The assessment prototypes, as they are based on a core set of eight cross-industry standard categories, form the basis for comparing scores and performances across a range of settings. Those involved in training programs and workforce preparation can use the challenging standards and assessments as a way of evaluating their own performance and improving the quality of training. More formally, it is also possible to score assessments resulting from different sites within an industry to see where training

might be more effective, to study such effective programs, and to disseminate information about how to better prepare workers for their industry.

For many involved in industry, it is not sufficient to measure employability skills or academic skills alone. These skills need to be integrated into specific industry contexts and to support specific industry needs. Without such an assurance, skill standards might seem like a "pie-in-the-sky" abstraction. However, both skill standards and the cross-industry categories of standards are used in ways that directly link to specific industry skills, those within the industry core, the concentration, and the specialty levels (see the NSSB framework in **Appendix A**). In fact, the assessment prototypes within this guide cut across these three levels. They adhere to a rigorous analysis of standards across industries, ensuring broad applicability and linkage to the core of many industries. At the same time, through the assessment development process, they are adapted to specific concentrations and specialities within an industry.

A vision for the workforce should not exist in isolation from education and a vision for education should not exist in isolation from the workforce. The environments are distinct yet they are both responding to changes in technology and methods of operation. The comprehensive vision proposed here, working in concert with the NSSB framework for industry skill standards, supports dialogue in the service of greater productivity both within classrooms and industry settings.

Chapter 3: Standards-Based Reform for Education and Industry

"Assessment based on industry-wide skill standards make clear to workers and employers what skills are needed. Standards provide a common language and framework for assessing competence. Within that framework, employers, individuals, educators, trainers, and managers will be working toward known benchmarks of competent performance."

— Fred Smith
Senior Vice President
American Electronics Association

Education and Industry as Partners

For many years, both in education and industry, professionals have relied on standards to help determine what students or workers should know and be able to do. Still, the visibility of the standards movement in America has increased significantly in the last fifteen years, largely because of concerns about the effectiveness of schools (*Nation at Risk*, 1983) and the competitiveness of American industries in the global marketplace (*America's Choice: High Skills or Low Wages*, 1990). These concerns have led people in both arenas to try to identify and define new ways to increase the quality of performance.

Standards have become a primary tool for reaching this goal of improving performance. They are, by definition, challenging directives that are linked to an ethic of quality. Giving all stakeholders access to information about what is valued, the standards provide impetus for self-evaluation and positive change on the part of individuals, work organizations, and public institutions.

Those in education and industry agree that setting high expectations for all students and workers is a fundamental step to improving their achievement levels. However, these same groups have had different goals and purposes as they have worked to develop standards. For employers, the difficulty comes when there is little transferability between what students learn in school and what they can do upon entering the workforce. Finding that they are currently using too many resources to train workers, industry aims to provide systematic and ongoing input into educational decision-making, beginning in the early grades, gaining in influence as students move into the later years of high school, and culminating in post-secondary occupational training programs. Increased input from employers about the scope and nature of the standards can increase the likelihood that students will leave school with knowledge and skills essential for successful performance in the workplace.

For educators, a goal has been to create a broad academic infrastructure from which students can make long-range decisions (MPR Associates, 1996). Most educators want employers to understand that students need curricula with sufficient depth and breadth to support further education and students' need for long-range skills, not only those contributing to entry-level employment. As technology and the workplace change, students also need an ability to adapt to new workplace environments and to seek out appropriate training for such an environment.

Educators also need broadly defined standards so that they can design curricula to meet the needs of all of their students. If standards are defined narrowly, "... it may become impossible to teach all aspects of an industry, or to contextualize necessary academic skills within an industry specialization" (MPR Associates, 1996). In order to make well-informed decisions about how to design industry-relevant instruction, educators need both access to challenging standards and direction on how to integrate applied learning, service learning, and career preparation into their existing curricula.

This chapter includes a broad definition of standards, types of standards used in industry and education settings, development and validation guidelines, and lessons learned through recent national and local efforts. The chapter closes with a summary of the work done by NSSAC to identify categories of cross-industry standards that are applicable across industries. This discussion lays the groundwork for subsequent chapters in this document on development of cross-industry assessment.

Defining Standards

A standard is an explicit statement that clearly defines the knowledge and skills and the level of performance expected of an individual in a given content or work area. As a set, standards represent consensus among stakeholders on what is most important for individuals in a field (of study or work) to know and be able to do.

Standards offer many potential advantages to various stakeholders. Overall, they provide a common language for workers, students, employers, educators, and community members to talk about performance expectations. As such, they afford the opportunity to build needed linkages and set common goals among the different stakeholders. Specific benefits to key stakeholder groups are described below (Far West Laboratory, 1995):

- workers know the underlying expectations for jobs and career development, enabling them to better meet employment criteria and increase their chances for mobility and advancement;
- students have clear goals for their educational and career preparation;

- employers have criteria to recruit, screen, place, evaluate, and promote employees more defensibly and efficiently; and
- educators have the necessary building blocks for designing high quality, focused programs and curricula consistent with the needs of key stakeholders.

Two different types of standards, *content standards* and *performance standards*, have been developed to meet these purposes. **Content standards** identify the areas of knowledge, understanding, and skills which are expected to be learned by individuals in key subject and career areas. They provide a basis from which educators, employers, union representatives, and other interested stakeholders can make informed decisions about educating or training their constituents, including *what* to emphasize in the curriculum or training program and *how* to approach instruction or training (see **Figure 1**).

Figure 1.
Example of a Content Standard (Health Care)

Legal Responsibilities*

Health care workers will understand their legal responsibility, limitations, and the implications of their actions within the health care delivery setting. They will perform their duties according to regulations, policies, laws, and legislated rights of clients.

The following may be included:

- Be aware of malpractice and liability issues
- Maintain client confidentiality
- Operate within scope of practice
- Comply with legal requirements for documentation

*From the National Health Care Skill Standards, 1995.

While content standards tell us what individuals should know, performance standards indicate how well we expect individuals to perform. **Performance standards** define and illustrate levels of expected accomplishment with respect to one or more content standards. Performance standards are used for a variety of purposes, including exemplification of content standards, as well as accountability and certification. They answer the question, "How good is good enough?" Performance, as well as content

standards, form the basis for designing and scoring assessments (see **Figure 2**). Often, performance levels are defined in three to six levels of proficiency (e.g., advanced, proficient, basic). To be made even more tangible, performance standards can be accompanied by specific examples of work that illustrate different levels of proficiency (WestEd, 1998).

Figure 2.
Example of a Generic Performance Standard — Used in an Assessment Aligned to Health Care Skill Standards

Written Scenario Part II: Overall (Holistic) Score			
Standard: Knowledge of ethical and legal responsibilities			
Limited	**Basic**	**Proficient**	**Advanced**
Shows little or no knowledge of ethical or legal responsibilities related to health care	*Shows gaps in knowledge of ethical or legal responsibilities related to health care*	*Shows adequate knowledge of ethical or legal responsibilities related to health care*	*Shows superior knowledge of ethical or legal responsibilities related to health care*

Serving different purposes, both content and performance standards are essential for building an assessment system. While content standards define the breadth and depth of knowledge and skills expected upon completion of the instructional program, performance standards define expected student outcomes and serve as the foundation of the scoring system used to evaluate student work. The combination of these two types of standards are important and necessary in designing and implementing assessment systems.

The following section discusses the range of contexts for standards usage — academic standards, employability standards, and industry skill standards — and the importance of integrating these standards to meet the needs of key stakeholders. Industry skill standards, obviously relevant and useable for industry, are often cross-walked or integrated with academic content standards (national, state, or local) for use in secondary and post-secondary settings. This process helps determine areas of overlap and importance for key stakeholders.

BEST COPY AVAILABLE

Descriptions of different contexts for standards — academic and industry — are provided, followed by a discussion of a general process for standards development and validation.

Academic Standards

Academic standards cover traditional subject matter areas such as mathematics, English/language arts, and science as well as areas such as geography and technology (see **Figure 3**). Written primarily for educators working in K-12 schools, these standards are used to develop challenging curricula and to link curricula to actual teaching practice. Most of the recently developed academic standards are intended to lead teachers toward designing instruction around major concepts in the discipline instead of around memorization of information pieces or practice of separate skills. As well, these standards are often benchmarked, showing what students at different grades or developmental levels should know and be able to do as they work toward understanding a concept.

Although academic standards are often developed at the state and local level, recent efforts by professional associations of educational stakeholders have been made to develop national academic standards for the major academic domain areas (e.g., science). For example, the National Council of Teachers of Mathematics (NCTM) has developed a set of national standards for mathematics education. Many states are using the national standards as a "launch pad" for their own development efforts.

Regardless of whether or not standards are developed at the local, state, or national level, they are public documents and thus likely to raise issues about what students should know. In order to help alleviate controversy over standards, it is important to include a variety of individual stakeholders and encourage broad consensus.

Figure 3.
Example of State Academic Standard

*** Reading: Comprehend a variety of printed materials**

Locate information and clarify meaning by skimming, scanning, close reading and other reading strategies.

Grade 10 Benchmark: Locate information and clarify meaning by using tables of contents, glossaries, indexes, headings, graphs, charts, diagrams, tables, and other reference sources.

*From the Oregon State Reading Standards, 1996.

BEST COPY AVAILABLE

Employability Skill Standards

Employability skill standards (also called generic *workplace readiness standards*) are the most general of workplace-related standards and cover skills and qualities that workers must have in order to learn and adapt to the demands of *any* job (e.g., communication, teamwork). In general, those who have identified employability skill standards share consensus about the abilities and skills needed in the workplace.

Employability standards are valuable to educators due to their general nature. Whereas occupational skill standards are usually narrowly defined, employability standards provide a more broadly-defined standard that can easily and explicitly be incorporated into the classroom. For example, *Teamwork* can be addressed in a variety of ways in the classroom and still retain its value to industry as well.

Perhaps the most well-known effort to identify employability skill standards is by the Secretary's Commission on Achieving Necessary Skills (SCANS, 1992). The SCANS project identified three foundation skills, with associated sub-skills, and five key competencies (see **Figure 4**). Included in these skills and competencies are basic academic skills that workers must have in order to succeed in most industries.

Figure 4.
SCANS Foundation Skills and Competencies

SCANS Foundation Skills	SCANS Competencies
Basic Skills • Reading • Writing • Arithmetic • Mathematics • Listening • Speaking	Resources • Allocates Time • Allocates Money • Allocates Materials and Facility Resources • Allocates Human Resources
Thinking Skills • Creative Thinking • Decision Making • Problem Solving • Seeing Things in the Mind's Eye • Knowing How to Learn • Reasoning	Interpersonal • Participates as a Member of a Team • Teaches Others • Serves Clients / Customers • Exercises Leadership • Negotiates • Works with Cultural Diversity
Personal Qualities • Responsibility • Self-Esteem • Sociability • Self-Management • Integrity / Honesty	Information • Acquires and Evaluates Information • Organizes and Maintains Information • Interprets and Communicates Information • Uses Computers to Process Information
	Systems • Understands Systems • Monitors and Corrects Performance • Improves and Designs Systems
	Technology • Selects Technology • Applies Technology to Task • Maintains and Troubleshoots Technology

Industry Skill Standards

Working jointly with the Department of Education to fund 22 pilot projects on skill standards across a range of industries, the Department of Labor openly encouraged the different industry groups to create standards to reflect their needs and economic environment and to target their standards primarily to entry-level workers within the industry or occupational cluster. As a result, industry skill standards were developed to address the knowledge and skills necessary to perform successfully in the workplace.

As of 1996, each of the 22 skill standard projects had created one or more sets of standards. Not surprisingly, many of these projects arrived at different conceptions of a skill standard depending on a range of factors, including (1) the needs of their industry, (2) how they defined the specific occupation or occupational area, and (3) the standards development and validation processes used (see **Figure 5** for an example of a skill standard that addresses content).

As a result, the skill standards looked very different across the 22 pilot projects. For example, some industries developed standards that addressed a single occupation (e.g., a sales associate for retail), while others developed a set of core standards and occupational cluster standards to capture the breadth of the industry (e.g., health care). In all cases, the pilot projects used development processes, and subsequently standard structures, that were useful and met their industry needs.

Figure 5.
Example of an Industry Skill Standard
(Hazardous Materials)

Job Function (A)*

Evaluate hazardous materials and hazardous waste sample data.

Supporting knowledge/skills:

- Perform mathematical calculations following existing formulas and reference materials
- Read and interpret blueprints, charts, curves, graphs, maps, plans, and spreadsheets from plotted and tabulated data
- Collect, tabulate, and assist in the evaluation of data, using appropriate techniques and technology such as:
  - calculators
  - computers
  - databases
  - graphics
  - spreadsheets
- Check laboratory and/or field sample analyses by comparing to regulatory limits

*From the Hazardous Materials Management Technology Skill Standards, 1995.

Standards Development/Adaptation and Validation Processes

In order to come to consensus on what students and workers should know and at what level they should perform requires input from key stakeholders as well as an extensive review and validation process.

There are four necessary steps in developing and validating standards. They include the following:

- conduct background research;
- produce draft standards using an inclusive stakeholder process;
- review and validate standards using multiple methods; and
- refine standards by pilot testing.

While local efforts may be unable to follow completely all the steps indicated, awareness of the goals of each step will help inform districts of the risks involved in shortcuts. External technical assistance may be required to accomplish some of these steps (WestEd, 1998). Each of the four steps is discussed in **Table 2**.

Table 2.
Standards Development/Validation Process

Standards Development/Validation Process	
Key Tasks	**Steps**
Conduct Background Research	• Identifying existing standards for adoption / adaptation reduces the risk of duplication. Many states, educational or professional agencies, and industries have established standards or competencies for what students and workers should know and be able to do.
Use an Inclusive Approach to Draft Standards	• Key stakeholders — policymakers, educators, employers, parents, students, employees — should be involved in all aspects of the development process. • An inclusive approach ensures that the needs of both educators and employers are met.
Review and Validate Standards	• A variety of methods can be employed to review and validate standards. One or more of the following review modes should be used: 1. *External Review Committee*—a large committee that includes representatives of key stakeholders. Committee members check for content appropriateness, clarity, and usefulness. These members should be independent from the group involved in drafting the standards. 2. *Surveys*—Mail surveys represent a cost-efficient means of securing widespread feedback and multiple perspectives on the relevance and importance of the draft standards. 3. *Focus Groups*—These groups are particularly useful for obtaining input from practitioners (e.g., teachers, employers, workers). Relative to other validation methods, focus groups provide some unique benefits, including the rich, in-depth information that emerges when participants respond to and build upon each other's different perspectives and thinking.
Refine Standards Through Pilot Testing	• It is important to ensure that diverse groups of students and workers participate in pilot testing application of standards (e.g., in curriculum, assessment). This ensures applicability and accessibility of standards to all learners and assists in providing for similar, high quality learning experiences for all cultural and ethnic groups.

NSSAC's Efforts

A central expectation of the skill standards "community" is that a common foundation of knowledge and skills needed across the majority of occupations within the economy will be identified (Hoppe & Wills, 1996). A major focus of NSSAC was to identify common, standard areas across the four participating industries that could be used as a basis for assessment development. The assumption here was that these standards would be essential for industry, applicable to secondary and post-secondary education, and linked to a high level of challenge and expectation.

NSSAC's Cross-Industry Analysis of Standards

NSSAC began identifying commonalities across the sets of industry skill standards from the four partner industries as a basis for creating common assessment prototypes (see Chapter 4). NSSAC partners represented a range of occupation specific standards to draw upon. The following examples illustrate the different types and structures of industry skill standards that had been developed by the four NSSAC project partners (only content standards are used as examples). The examples all reflect the common category of teamwork.

WestEd — Health Care

The National Health Care Skill Standards Project (NHCSSP), created broad and generic standards for the health care industry to accommodate both the diversity of occupations and the changing nature of the industry. More specifically, the standards included both a core level and four sets of more specific occupational cluster standards. Given that the health care industry is well-organized with multiple levels of existing certifications (Grubb, 1996), it was necessary that the standards provide a broad base within this industry context. The standards are designed to apply primarily to entry-level and technical-level workers (**see Figure 6**). Each standard statement consists of these parts: (1) a brief title, (2) a description of the knowledge and skill, and (3) specific points or examples intended to clarify the standards.

Figure 6.
Example of a Health Care Skill Standard

Teamwork*

Health care workers will understand the role and responsibilities of individual members as part of the health care team, including their ability to promote the delivery of quality health care. They will interact effectively and sensitively with all members of the health care team.

The following may be included:

- Practice team membership skills, such as cooperation, leadership, and anticipation of the needs of coworkers
- Respect cultural and religious differences of team members
- Interact with others consistent with the health care team structure and lines of authority
- Manage conflict within the workplace through consideration of others' points of view
- Respect interdisciplinary differences among team members

*From the National Health Care Skill Standards (Core), 1995.

National Retail Federation — Retail

In contrast to the Health Care Skill Standards, the NRF Skill Standards focus on *one* occupational area, Professional Sales Associate. There are a total of six skill modules that target this occupation. They offer a flexible framework for standards use, with certain modules more relevant to some sales situations than others. The professional sales associate represents a large percentage of the workforce and a large percentage of first-time employment situations. Retail workers are often younger and less likely to have post-secondary degrees than workers in other industries. As well, retail work tends to involve part-time jobs (Bailey & Bernhardt, 1996) but provides opportunities to learn the types of customer service and sales skills that are useful in many industries (see **Figure 7**).

Figure 7.
Example of a Retail Skill Standard

Module Six: Work as Part of a Department/Store Team*

6.1 Support Co-Workers
6.1.1 Share ideas and information about selling, marketing, and products

How might the task be demonstrated:

- Share information fully and in a timely manner.
- Contribute experiences and knowledge of products with fellow associates.

*From the National Retail Federation Skill Standards, 1995.

Human Services Research Institute — Human Services

The Community Support Provider Skill Standards also focus on a single occupational area, the Direct Service Worker. The standards are organized into twelve standard units, each representing a broad competency area (see **Figure 8**). Specific functions for each area are broken into two to five skill standards with each skill standard linked to one or more realistic sample activities. Each sample activity is followed by one or more performance indicators, consisting of observable worker behaviors, client reports, and worker self-reports, which provide a basis for measuring performance.

A majority of direct service workers now work in community-based organizations demanding different sets of skills from those needed in institutional settings. Excluding institution-based service workers, the standards reflect industry direction and a philosophical shift toward community-based services.

Figure 8.
Example of a Human Services Skill Standard

Competency Area 4: Community and Service Networking*

The community support human service practitioner should be knowledgeable about the formal and informal supports available in his or her community and skilled in assisting the participant to identify and gain access to such supports.

Skill Standard C — the competent CSHSP ensures participant access to needed community resources coordinating supports across agencies.

Activity: the competent CSHSP collaborates and shares information with staff in external service and support organizations.

*From the Human Services Research Institute Skill Standards, 1995.

American Electronics Association — Electronics

The American Electronics Association (AEA) Skill Standards consist of five components, four sets of skill standards for different occupational areas and a framework for foundation skills and workplace competencies (see **Figure 9**). The foundation skills and competencies are derived from SCANS. AEA supplemented the SCANS categories and tailored SCANS definitions, with changes related primarily to technological and business-specific skills and competencies.

Designed to be applied primarily to non-baccalaureate level workers and to describe work at a mastery level, the AEA Skill Standards represent four different occupational areas, each consisting of five to seven critical functions or "competency modules." Two to five activities define each critical function and each activity has from three to nine performance indicators which offer ways of assessing whether the activity has been performed competently. The system is designed to enhance flexibility so that demonstrating competency in a critical function in one occupational area is likely to be transferable to other occupations and fields.

Figure 9.
Example of an Electronics Skill Standard

Critical Function: Initiate and sustain communication processes and procedures*

Key Activities: Create and enhance effective, productive relationships within the work group

How do we know when each key activity is performed well?

- Constructive feedback and active support are sought from and provided to individuals within the work group and to the work group as a whole.
- Co-workers and colleagues are treated with professionalism and respect at all times.
- Disagreements, conflicts, and grievances are settled in a positive and timely manner according to company procedures.
- Individual and work group roles and expectations are defined.

*From the American Electronics Association Skill Standards, 1994.

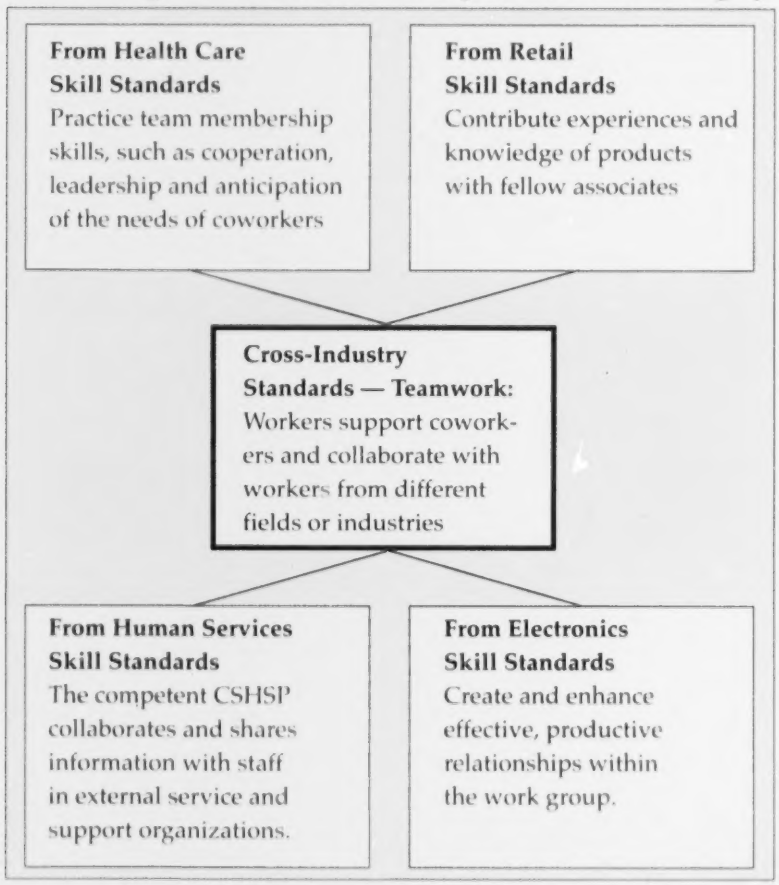
NSSAC Cross-Industry Categories of Standards

Although NSSAC's four sets of industry skill standards differed in structure, language, and level of specificity, commonalities seemed readily apparent (e.g., the continuing thread of teamwork in the example standards). Systematic identification of these common categories was accomplished by using the SCANS (1992) skills and competencies as an external referent to help establish a basis for analysis, as previously discussed. SCANS is a well-defined and accepted set of general workplace skills and competencies that are appropriate to use as a referent in thinking about different standard sets. This technique is consistent with Center for Workforce Development's *In Search of Commonalities* (1996), as described previously. The common categories thus identified are shown below. These cross-industry categories were generated from sets of validated standards but do not in themselves represent validated standards.

An example of a category identified as common across all skill standards is teamwork. Each set of standards addresses teamwork in some capacity, although each approached the concept differently. For example, teamwork is an explicit standard in

health care (see **Figure 6**), while in electronics, the teamwork concept is included within a communication standard (see **Figure 9**). **Figure 10** demonstrates the teamwork component for each of the four sets of skill standards, as well as the teamwork cross-industry category.

Figure 10.
Excerpts from Skill Standards (Teamwork) and Resulting Teamwork Cross-Industry Skill Standard Category


From Health Care Skill Standards
Practice team membership skills, such as cooperation, leadership and anticipation of the needs of coworkers
From Retail Skill Standards
Contribute experiences and knowiedge of products with fellow associates
Cross-Industry Standards — Teamwork:
Workers support coworkers and collaborate with workers from different fields or industries
From Human Services Skill Standards
The competent CSHSP collaborates and shares information with staff in external service and support organizations.
From Electronics Skill Standards
Create and enhance effective, productive relationships within the work group.

In all, the cross-industry analysis yielded eight inductively-generated categories (see **Figure 11** below) with each representing a major area of focus across the four industries. These categories form a flexible framework, allowing industries to design and adapt assessment methods to meet their specific needs and context. The design and results of the NSSAC cross-industry analysis of standards is summarized in a recent research report (Schwager, 1997).

Figure 11.
The Eight Cross-Industry Categories or Competency Areas

- *Client Orientation*: Workers are sensitive to customer/client needs and adapt interactions, services, and products to best match these needs. New products, services, and markets are created as workers perceive the demand and innovate to best meet it.
- *Teamwork*: Workers support coworkers and collaborate with workers from different fields or industries. For example, human service workers may facilitate services and collaborate with client's employer, school personnel, community service agencies, and health care agencies.
- *Evaluating and Interpreting Information*: Workers assess contexts and clients/customers and use this information to assist with decision making and goal setting. Feedback on the quality of industry services is used to improve quality and maintain a competitive edge.
- *Awareness of Safety and Security Practices*: Workers understand human, legal, and civil rights involved in appropriate practice in their field and work to protect the interests of their clients/customers or the company/organization.
- *Applying Technology*: Workers use technology across a variety of tasks both electronic and mechanical.
- *Organizing Information*: Workers research, develop, and maintain databases or files of information about the customers/clients and about available services, products, and resources.
- *Managing Resources*: Workers manage or assist clients with managing schedules, equipment, therapies or services, and materials.
- *Professional Orientation*: Workers establish connections to a larger professional context. They complete necessary training, engage in continued professional development, and educate clients/customers and the community about relevant issues.

The eight cross-industry categories are not fundamentally different from selected SCANS and general workplace readiness skills. In fact, findings from the NSSAC analyses are supported by the results of a broad-based study identifying critical behaviors in a variety of occupations. The *National Job Analysis Study* (1995) had workers identify and rank work-related behaviors across a broad representative sample of occupations. Five of the ten most important behaviors — listening to the concerns of clients/customers and responding; using a computer to locate, process, or communicate information; providing information to people; judging the importance, quality, and accuracy of information; and listening to instructions from or concerns of supervisors or co-workers and responding — directly relate to the SCANS-like competencies and foundation skills identified as common to the four industries.

These categories are certainly not comprehensive. Other common skill areas may exist among the industry standard sets, and other standard sets may lead to the development of additional or different categories. At present, the cross-industry categories are being used as a basis for developing assessment methods viable both across industries and within secondary and post-secondary education settings.

Both industry representatives and educators have valid concerns, and the challenge has been to align these different perspectives into an integrated system. While standards can provide a common language and focus as an essential part of an occupational assessment and certification system, "...it is not standards per se, but ongoing communication in and among all levels of the system about what is expected" from students and workers that will drive innovation and keep the system dynamic and up-to-date (Jenkins, 1994). It is this ongoing and unchartered communication between education and industry that provides hope for a flexible, usable, and effective system.

Chapter 4: Qualities of an Effective Occupational Assessment and Certification Program

An industry association, state, or school district can begin to address issues of assessment and certification after it has developed or adopted standards. This chapter describes the qualities of an effective program and the steps that should be taken to ensure that these qualities will be achieved. It is assumed that the purpose of certification is to qualify individuals for specific occupations or more general workplace skills and behaviors. The first section delineates key criteria to consider in developing a high quality assessment and certification system; the next section offers examples of assessment designs for different certification programs.

"We look forward to a future where classroom education and workplace training are highly relevant; where students and workers have a realistic view of what they know now and what remains to be mastered to meet their goals; and where competence is recognized and valued. Progressive assessment strategies aligned with competency based standards are the path to that future."

— Marianne Taylor
Project Director
Human Services
Research Institute

Assessment and Certification for the Workplace

Assessments serve an *accountability* function, that is, they ensure service or product quality by holding workers accountable for their competence and productivity. Assessment and certification programs are typically designed to lend credibility to and support the training and professional growth of those working within an occupation. *Certification* involves an individual going through an assessment process to show that he or she has met a pre-established set of qualifications, usually a voluntary process administered by a nongovernmental agency (Smith, 1996). In contrast to certification programs, *licensure* and *credentialing* programs are mandatory and governmental. Such programs demand that individuals demonstrate a minimum level of competence in order to work within a profession.

Ensuring the Technical Quality of an Assessment and Certification Program

Just as it is essential to involve industry personnel, it is also important to involve assessment experts, or psychometricians, in the design, development, and implementation of an assessment and certification program. Such an expert can help evaluate the technical quality of the program to help ensure its legal defensibility. This section is organized around seven criteria often used to evaluate the technical quality of an industry skill standards-based assessment and certification system. Each criterion will be described in turn.

Criteria to evaluate the technical quality of an assessment and certification system:

- Is the assessment aligned to challenging standards?
- Does the assessment incorporate multiple measures?
- Is the assessment valid and does it reflect what is valued in the workplace?
- Is the assessment reliable and free from unacceptable levels of measurement error?
- Are the results comparable across different sites?
- Is the assessment equitable and fair?
- Is the assessment practical and cost effective?

Is the Assessment Aligned to Challenging Standards?

Assessments should be based on challenging standards of which both the employer or instructor and the employee or student are aware. Challenging standards, as shown in Chapter 3, broadly define the complex problem-solving skills and behavioral characteristics needed in the workplace. Because these standards are applicable across a wide range of settings, they support worker mobility and the ability to adapt to changing conditions in the workplace.

A considerable amount of research has shown that real jobs across a range of industries, whether entry-level or more advanced, often involve significant uncertainty and variety in the nature of the tasks performed (Wirt, 1993). For this reason, an assessment system linked to challenging standards provides both a detailed picture of a worker's skill level, as well as a more valid picture of worker potential and competence. Employers can then select, train, or promote an employee who performs beyond minimum competence and meets those higher standards that will ensure product or service quality.

Such challenging standards provide a common basis from which assessments can be developed, performances can be evaluated, and decisions can be made about worker competence. It is essential to check throughout the development process to make sure that there is alignment among the standards, the assessment tasks, and the criteria being developed to score the assessments.

Does the Assessment Incorporate Multiple Measures?

Assessments are often composed of different kinds of tasks, such as traditional paper and pencil and performance-based tasks, each of which provides a different piece of the overall picture of a student's or worker's performance. It is important to employ a variety of assessment tasks, as each kind taps into different kinds of knowledge and demands a different cognitive process for arriving at a response (Resnick and Resnick, 1992, among others). The most valid and complete pictures of performance are gained through assessments that use multiple measures and that broadly tap into the valued knowledge, skills, and abilities as defined in the content standards.

This approach is referred to as *triangulation*, or the accumulation of multiple sources of evidence as the basis for sound decision-making (Miles & Huberman, 1994). More formally, triangulation is defined as "the attempt to obtain more valid results about one aspect of performance by using multiple sources of data, multiple methods of collecting data, and/or multiple interpretations of the same data" (Wheeler and Haertel, 1993).

An overview of different kinds of assessment tasks is provided below.

A multiple-choice item presents respondents with a highly-structured question, accompanied by four or five possible answer choices. It is a form of "selected response" assessment because respondents are asked to choose the correct or best answer from the options presented. Multiple-choice assessments strive for fairness and consistency by (1) using a large number of questions that can be easily administered and machine scored, (2) "decontextualizing" questions so that differences in prior knowledge are minimized, and (3) administering exams under standardized conditions to eliminate any variability in time allowed or directions given (Wirt, 1993). The large number of items (e.g., 30-50 items for a 60-minute test) in a multiple-choice test also can result in broad coverage of relevant content standards.

The above list of characteristics clearly are advantageous for high-stakes assessment purposes. At the same time, multiple-choice tests often lack the credibility of more realistic measures of worker competence as they do not adequately reflect the complex, integrated skills that are often required in the workplace. In fact, it may be difficult to create demand for a multiple-choice

type of assessment and certification program in a field "in which employers view paper and pencil testing as a limited measure of competence" (Electronic Industry Foundation, 1996). In short, the cognitive demands of such a test are often not perceived as parallel to the cognitive demands of the workplace.

In recent years, many involved in assessment development have experimented with new ways of writing multiple-choice items so that they more accurately reflect complex problem-solving skills and are contextualized within an industry setting or an academic discipline. These strategies have increased the relevancy and credibility of such items.

For example, the following two questions (taken from sample assessment items in the career area of Food Services and Hospitality) illustrate how multiple-choice questions can be designed to measure simple recall of memorized information as well as deeper levels of understanding involving the application of knowledge (see **Figure 12**). In this case, respondents apply knowledge of the *Food Guide Pyramid* to select the most balanced and healthy menu from several choices.

Figure 12.
Examples of Recall and Application
Multiple-Choice Items

Sample item assessing recall of knowledge:

According to the Food Guide Pyramid, how many servings from the "bread, cereal, rice and pasta " group should an individual eat daily?

A. 2-3 servings
B. 2-4 servings
C. 3-5 servings*
D. 6-11 servings

Sample item assessing application of knowledge:

Which menu is most healthy and includes foods from each major food group in the Food Guide Pyramid?

A. Hamburger with lettuce and tomato
   Potato chips
   Carrot sticks
   Sliced watermelon
   Diet soda

B. Spaghetti with meatballs
   Garlic bread
   Fruit salad
   Ice cream
   Iced tea

C. Turkey breast sandwich*
   Mixed green salad
   Pretzels
   Frozen yogurt
   Apple cider

D. Stir-fry chicken with peanuts
   Rice
   Mixed vegetables
   Almond cookie
   Milk

* correct response

Source: WestEd, 1998.

Multiple-choice tests have an other potential disadvantage — they are often seen as separate from and external to the ongoing process of learning on the job (being instructed by a trainer or being evaluated by a supervisor). The most relevant assessments are often ones that are embedded in the workplace and part of everyday workplace protocol (Wirt, 1993).

Constructed response items, such as written scenarios and essay tests, require that examinees prepare a written response to a prompt. These items are typically on-demand (i.e., administered on a specified date, under time constraints) and designed to reveal what a student or worker knows at a given point in time about a given subject. In such *open-ended* items, there are multiple pathways to providing a response and typically more than one right answer (typically, multiple-choice items have only one right answer). Those who score responses to such items are trained to look for specific kinds of evidence — such as evidence of conceptual understanding, evidence of a workable solution to the problem, and evidence of the ability to communicate knowledge and understanding in a coherent way (see **Figure 13** for an example from Health Care).

Figure 13.
Example of a Constructed-Response Item

Construct a temperature-pulse-respiration chart recording the vital signs for four patients you monitored during a night shift. Describe how you would report the status of each patient, particularly any variations in vital signs you would want to draw to the attention of the charge nurse.

At midnight vitals were:			
Marion Jones	98.6°F	88	20
Peter Smith	97.4°F	96	16
Lila Carrie	98.2°F	72	14
Adrian Shiff	101.8°F	104	22
At 4 A.M. vitals were:			
Marion Jones *(temperature taken by axilla)*	99.6°F	92	16
Peter Smith	98.2°F	78	18
Lila Carrie	98.2°F	74	20
Adrian Shiff *(temperature taken rectally)*	102.8°F	112	24
At 8 A.M. vitals were:			
Marion Jones	98.8°F	80	16
Peter Smith	99.0°F	86	18
Lila Carrie	97.8°F	88	16
Adrian Shiff	101.6°F	96	22

Source: adapted from Schniedman, Lambert, & Wander, 1989.

Performance-based assessment tasks such as computer simulations, case studies, long-term projects, and portfolios more closely reflect "real-life" situations facing workers.[1] Such items, or collections of items, often are more authentic and engaging, requiring a longer-term investment of time and energy and showing more clearly the quality of one's product or performance (Wiggins, 1993). Such tasks also make greater demands on academic foundation and problem-solving skills than traditional paper and pencil assessments (Rabinowitz, 1997).

Performance-based assessments also allow for "stronger links among assessments, standards, and instructional/training goals"

[1] Many assessment specialists consider constructed response items as a subset of performance-based assessment.

(Rabinowitz, 1997). They demand direct evidence of the ability to actually perform integrated, complex skills; they can be embedded in workplace contexts; and they often have high credibility.

Drawbacks of performance-based assessments include their lower reliability, and greater development, administration, scoring, and training costs. In fact, in some situations performance-based assessments have been cost prohibitive (Institute for Educational Leadership, 1996).

Other workplace assessment strategies include interviews, competency checklists, direct observations of work, and client/customer feedback. Each of these strategies can be a viable part of an assessment and certification program.

Is the Assessment Valid and Does it Reflect What is Valued in the Workplace?

When designed with a clear purpose, an occupational assessment and certification system will more likely be seen and demonstrated as *valid*. Validity refers to "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test *validation* is the process of accumulating evidence to support such inferences" (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985). There are numerous test validation processes used to find out what an assessment actually measures and what scores on the assessment actually mean, including but not limited to *content, predictive,* and *construct validation*. Though these are formal processes requiring some technical expertise, validation involves making a concerted effort to learn what the assessment tasks actually measure, ensuring that they capture important aspects of job performance and avoid factors that are unrelated to job performance.

Content Validation. Content validation involves demonstrating that the content of the assessment adequately *samples* the domain of knowledge, skills, and abilities related to the occupation (Institute for Educational Leadership, 1996). As it would be infeasible to assess all aspects of a test-taker's relevant knowledge, skills, and abilities, *sampling* makes it possible to draw conclusions about the whole through evaluating representative parts. Such sampling, however, must be deemed truly representative of the domain in question.

An informal aspect of content validity is *face validity*; an assessment that is "valid on its face" is perceived as acceptable and legitimate by key stakeholders and the general public (Wheeler and Haertel, 1993). Face validity is not a formal measurement concept, but it is important for all stakeholders to have the impression that the assessments are doing what they are designed to do. One would not want to give an automobile driving test to someone who is trying to be certified as an airline pilot; this would not have face validity (even if it had other more formal kinds of validity). The occupational assessment and certification system must measure the knowledge, skills, and abilities that are seen as relevant to the occupation.

Predictive Validation. Though most credentialing and certification exams are currently evaluated for content representativeness, assessment specialists also investigate the extent to which the performance on the assessment predicts performance on another measure. For example, an assessment designed to predict future performance may be given to persons applying for entrance into a training program. This assessment is not used to make decisions about admission into the training program; instead, the results are saved and compared to how each of the people actually perform in the training program. Through this predictive validation process, the relationship between scores on the assessment and later performance is determined (Institute for Educational Leadership, 1996).

Licensure and employment tests must demonstrate content and predictive validity — that is, they "must either contain content minimally necessary for successful job performance or accurately predict those who will be successful and unsuccessful in a job" (Phillips, 1995). This requires a formal study relating job requirements or performance to the knowledge, skills, and abilities assessed.

Construct Validation. Construct validation evaluates how well an assessment measures *constructs* or attributes of performance that are of interest. Instead of being seen as discrete parts of the knowledge or skill domain, constructs are complex psychological and behavioral characteristics that are related to job performance, such as critical thinking or problem solving. The patterns of results of performance on new assessments is compared with the pattern of results of performance on other high quality assessments. When similar assessments produce similar results and dissimilar assessments produce dissimilar results, greater

confidence can be placed in the new assessments' ability to measure the construct of interest (Stecher, Rahn, Ruby, Alt, Robyn, with Ward, 1997).

Many measurement experts now agree that construct validity is the unifying force or overarching strategy for validation studies and recommend examining *consequential validity*, or the effects of the assessment on practice, as well (Messick, 1989; Linn, 1994).

Is the Assessment Reliable and Free from Measurement Error?

To be reliable, an assessment must produce scores that reflect the level of competence or skill level of a worker as accurately as possible. However, scores resulting from any assessment contain some degree of error. Perfect reliability (i.e., a reliability index of 1.0) is a theoretical concept and never actually achieved in practice. Instead, a reliability of .80 or above is considered an acceptable value for certification programs (Knapp, 1995).

Measurement errors that reduce reliability can result from (1) inadequate or inappropriate selection of specific tasks from the domain of relevant knowledge, skills, and abilities (*content domain sampling*), (2) disagreement among people scoring the assessments (*inter-rater reliability*), or (3) differences in the conditions under which the assessment was conducted (*standardized operations*). Each of these sources of error will be described in turn.

Content Domain Sampling. An assessment cannot measure everything contained in the standards — there is not enough time or resources to construct and take such a comprehensive examination. Instead, assessments typically *sample the domain*, or evaluate one's knowledge of representative parts of the whole.

In general, a longer assessment will include more items, will sample more of the content domain, and will be more reliable than a shorter test containing fewer items. The longer test will therefore provide a more solid basis for making judgments about the examinee's overall abilities.

However, the kinds of assessment items used also have a direct impact on reliability. Performance-based tasks require an examinee to take the time to prepare a more in-depth response. For this reason, in a one-hour test period, one can perhaps administer one to three performance-based or constructed-response tasks

when it might be possible to administer 50 multiple-choice tasks in the same amount of time. For this reason, multiple-choice tests cover more of the domain, albeit with less depth, and have traditionally been deemed more technically reliable given the same amount of time for administration than performance-based assessments.

The reliability of performance-based and constructed-response tasks can be low partly because an examinee's performance on one task often cannot be *generalized* or show a significant relationship to how the person performs on other tasks. This might be due to unfamiliarity with specialized equipment or technology used in a performance task.

Thus, examinees might "get lucky" and be asked to perform a task that is familiar to them. On the other hand, examinees might not be so lucky and might be asked to provide an in-depth response about a small part of the domain about which they know very little. For this reason, it is important to use performance-based assessments primarily in situations when the performance skill itself is a significant part of what is being measured (Phillips, 1993). Similarly, constructed-response tasks are best used when the performance skill of communicating a response is an important part of what is being measured.

Inter-Rater Reliability. When a score is assigned through the use of rater judgment, inter-rater agreement is an important component of assessment reliability. Often, constructed-response or performance-based tasks are scored by multiple raters and the score assigned by two or more raters to an individual response are compared to determine the agreement of the judgments being made. If scorers are routinely in high agreement, the assessment will have high inter-rater reliability and be relatively free of measurement error associated with scorers. Conversely, if scorers routinely disagree about the quality of performance, then the assessment will have low inter-rater reliability, indicating more measurement error.

When designing the scoring of performance-based and constructed-response tasks, it is important to document how scorers were selected and trained and what scoring procedures were used. (See Chapter 6: Scoring and Reporting for more details.)

Standardized Operations. Some measurement error can be avoided if an assessment is administered in a standardized

manner — that is, if each candidate receives identical directions, the same amount of time to complete each assessment item, and the same or comparable items. Also, if specialized equipment is used, such as computers, it is important to give examinees enough time to become familiar with the equipment prior to the assessment. In such situations, one would send identical sample material to each examinee prior to administration of the test (Phillips, 1993).

In addition to all candidates receiving clear and identical guidelines, assessment personnel must be trained to administer the assessments in a standardized format and to protect the security of assessment materials. The higher the stakes for the assessment, the more secure it must remain. Administration and scoring site personnel are to make sure that no secure items leave the site; any problems or irregularities should be reported to the governing association.

It is also standard protocol to protect the confidentiality of all candidates, including the scores they receive on an assessment. Standardized procedures for providing feedback to candidates is necessary to protect confidentiality and to ensure that results are used appropriately.

Are Scores Comparable Across Different Sites?

For assessment tasks that are embedded in the workplace and collected on an ongoing basis, one of the keys is to find ways to develop, monitor, and assure the consistency and comparability of judgments across widespread sites or parts of the system (Wirt, 1993). One strategy that can be used to make assessments more comparable to each other involves using an assessment *prototype*, or template, to guide development.

A prototype or template serves the purpose of ensuring that an assessment task contains all of the necessary information presented in an identical way. For examinees, using such a template means they can gain familiarity with the process of responding to such a task and know what to do in a potentially stressful testing situation. For the assessment developer, the prototype provides a template for development and review of tasks. In these situations, valuable time and resources are not lost reinventing the format; instead, time can be devoted to developing appropriate, specific content.

A prototype can be format guidelines or it can be more developed and linked to a specific kind of task, problem, or situation. One can adapt such a prototype to multiple industry contexts, situations, and settings.

In the case of portfolios, the prototype ensures that each student or worker submits the same kinds of entries, even though they may have different foci. By defining the criteria for what should be included in the items, one can later construct scoring criteria that can be consistently applied across all portfolios. Using such a prototype will increase the likelihood that two overall performances are comparable, even if they originate from different sites.

Equating. In order to maintain security within high stakes testing situations, it is often necessary to have more than one version of a test. Though the specific items on the test form differ, the goal has been to make sure that each form measures the same skills. *Equating* is then the process of adjusting scores from different versions of a test to account for differences in their overall difficulty level.

For multiple-choice tests, equating procedures are well established and technically defensible. However, performance-based tasks pose problems for equating because the tasks are multidimensional, unlike multiple-choice items (Holmes, 1985). It is much harder to design parallel performance-based tasks that measure the exact same thing. By using a prototype as the basis for development, one is more likely able to equate test items from different versions of the same test. For example, two written scenarios designed for use on a dental assistant exam could pose different situations or applications that measure parallel knowledge, skills, and abilities within that industry context.

Is the Assessment Equitable and Fair?

One way to help ensure the fairness of an assessment and certification program is to include representatives from all segments of the workforce in terms of gender, race, ethnicity, region, and industry settings. Such representation is essential at each stage — that is, during development of standards, the assessment and certification program, the specific assessment tasks, and the scoring criteria. Equitable involvement at each stage can help to prevent problems related to bias and inaccessibility of the assessment to some populations.

After assessment tasks and scoring criteria have been developed, it is also important to conduct a *fairness review* in which informed experts evaluate the material for potential adverse impact or bias. For example, reviewers evaluate both the use of language to make sure that it is sensitive and appropriate and the content to make sure that people from one subgroup are not more likely to perform well than those from another subgroup. Reviews try to identify situations which can be misunderstood because of language deficiencies, personal or cultural reactions to the task, unfair knowledge and familiarity with equipment, or other variables (Phillips, 1993).

To be in accordance with civil rights law, an assessment must measure qualities linked to job performance and not to personal qualities of an individual unrelated to job performance. In fact, the Civil Rights Act of 1991 considers any employment practice unlawful if:

> a complaining party established that [an employer] uses an employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and...fails to demonstrate that the challenged practice is job related for the position in question consistent with business necessity...
>
> (as quoted in Guttman, 1997).

A fairness review helps to ensure that the assessment is linked to job performance only and not to unrelated factors. Basing the development of assessments on validated standards linked to specific industry contexts also is important in ensuring job-relatedness.

Even when involving all constituents and employing a fairness review, an assessment and certification program is still obligated to monitor assessment results. Substantially lower performance by historically disadvantaged groups relative to others may or may not signal bias; such trends should be furthered investigated to ensure absence of bias.

Assessments developed employing the above-described strategies, when used as the basis for hiring decisions, are more likely to withstand challenges in the courts because they have documented job-relatedness. At the same time, the threat of adverse impact on certain groups has been a real challenge in high-stakes

testing. Another important strategy for alleviating the problem has less to do with assessment design and review: providing equal access to standards, education, and training.

Is the Assessment Practical and Cost Effective?

Developing an assessment and certification program can be costly. Accepted practices include developing assessment tasks, conducting validity and reliability studies, marketing, and programming and systems development. Typical costs for a multiple-choice type of assessment and certification system can approach $250,000 for development alone. A program certifying 10,000 candidates per year could have costs near $1,000,000 for administration and updating assessments (for an overview of program costs, see Knapp, 1995; for detailed information of costs involved in existing certification program, see Electronic Industries Foundation, 1996). A performance-based assessment and certification system can, at a minimum, double program costs, with much of the added expense going to hiring and training people to score candidates' work in a reliable and legally defensible manner. In fact, some studies of student assessment systems estimate the cost differential as much higher with performance-based assessments often costing five to ten times higher, and more (Stetcher, et. al., 1997).

Establishing an Assessment and Certification Program

Attention and care in defining a program is as important in its successful operation as ensuring that it is of high technical quality. Choosing the most appropriate assessment methods will have a direct impact on the credibility and quality of the program. A well-designed and legally defensible assessment process can take many forms, depending on its intended purpose.

All of the factors discussed previously that relate to ensuring technical quality need to be considered in the design of the assessment or certification procedure itself. With the assistance of assessment experts, a test blueprint can be developed that addresses the kind, number, and cognitive complexity of items, the time frame for implementation, and the degree to which the work is independent. At all times, programs need to be designed that are relevant to and meet industry needs. For example, although all programs may need to assess problem-solving and flexibility, these skills may need to be established more for some types of work than for others. We can expect that jobs that utilize

detailed protocols and uniform procedures may have less need to establish the ability to generate new solutions or respond to emergency situations than less well-defined work.

These kinds of considerations figure prominently in blueprint design. For example, the assessment design for existing occupational assessment and certification programs linked to the National Institute for Automotive Service Excellence calls for 40 to 80 multiple-choice questions for each test, with questions stressing real-work situations and problems. Other types of assessment designs may stress practical projects, such as the Federal Aviation Administration certification programs, or documentation of interactive episodes through videotapes, such as the portfolio component used as part of the assessment design for certification by the National Board of Professional Teaching Standards (see **Figure 14** for full assessment designs).

Figure 14.
Two Sample Assessment Designs

Assessment design for the Federal Aviation Certification programs:

- Certification involves written, practical, and oral tests.
- Each written test includes 75 objective multiple-choice questions.
- Practical tests involve projects assigned by an FAA flight standards inspector or FAA-designated mechanic examiner.
- Oral testing includes questions related to the practical projects and other questions used to determine overall competence.

Assessment design for certification by the National Board of Professional Teaching Standards:

- Portfolio containing five separate entries to be developed over the course of the school year. The portfolio contains videotapes of classroom practice, samples of student work, analysis of these products, and other evidence.
- A six-hour assessment at an assessment center in which candidates respond on-demand to four separate exercises.

Chapter 5: NSSAC's Cross-Industry Assessment Prototypes

For NSSAC, assessment development and pilot testing focused on assessment methods other than standardized, multiple-choice items. Use of multiple-choice assessment is already well-established as a cost efficient, legally defensible, and reliable way to assess and certify proficiency on targeted standards. The limitations of this methodology are also well known, namely, that such tests do not capture adequately what individuals think, do, and actually produce.

Performance-based assessment, in contrast, is increasing in popularity. Despite growing support, the usability (e.g., technical soundness, cost efficiency, standardization) of such methodologies for large-scale or high-stakes assessment purposes, such as certification, remains limited.

"Careful use of student assessments promotes the partnership of vocational education with industry by ensuring a smooth transition from the classroom to the job site."

— Pamela Cooper
Lead Teacher, Dental Assistant Program
Grossmont Health Occupations Center
Santee, California

A major NSSAC objective was to explore the viability of selected performance-based methods for skill standards-based assessment and certification across various industries. Our efforts focused on two different performance-based assessment methods — the *written scenario* and *portfolio*. The written scenario is an on-demand assessment, designed to be administered in a set amount of time and to measure what a respondent knows at a given point in time. The portfolio, on the other hand, is a *cumulative* assessment. It is designed as an on-going assessment to be completed over a period of time as a record of growth, progress, and overall level of performance.

In particular, NSSAC sought to develop *prototypes* for both written scenarios and portfolios — that is, templates or models having general requirements — that can be adapted and applied to specific industries, such as retail or health care. Using such a prototype, those working within specific industry contexts can efficiently create appropriate assessment tasks, likely to be comparable across settings within the industry.

NSSAC examined *both* written scenarios and portfolios for two additional reasons: (1) to provide different but complementary performance-based assessments to help create a strong basis for communication among educators and industry representatives; and (2) to provide an avenue for more comprehensive coverage of the cross-industry categories (see **Figure 15** for a list of NSSAC categories; for more detail about these categories and how they were derived, see **Chapter 3**). For example, *Applying Technology*, one of the eight cross-industry categories, is difficult to measure in a written scenario because of the necessary hands-on

component. Ample evidence of one's ability to use technology, however, can be documented in a portfolio. On the other hand, evaluating effective *Client Orientation* techniques within an industry context is effective using a scenario format because having a worker explain in writing how he or she would handle a client-centered situation or problem taps important standard-based knowledge and skills.

Figure 15.
The Eight NSSAC Cross-Industry Categories

- Client Orientation
- Teamwork
- Evaluating and Interpreting Information
- Awareness of Safety and Security Practices
- Applying Technology
- Organizing Information
- Managing Resources
- Professional Orientation

Some cross-industry categories can be measured using both a written scenario and a portfolio. In these cases, the combination of the two assessment methods tap into different aspects of performance and work to provide an accurate picture of worker competence.

Following is an in-depth description of both assessment methods, including an explanation of the development process NSSAC employed and resulting prototypes that can be applied across industry contexts.

The Written Scenario

The written scenario assessment requires individuals to respond in writing to a scenario depicting a complex and realistic problem that workers confront in an occupational context. The respondent must read the scenario, think about possible solutions, organize his or her thoughts, and propose a solution in writing, all within a given time period. The response must demonstrate a respondent's ability to apply important knowledge and skills to real world situations related to the targeted occupation or industry (see **Figure 16**).

BEST COPY AVAILABLE

Figure 16.
Overview of a Written Scenario

TITLE ⇨	**Facilitating Services**
PROMPT ⇨	You are a community case manager and are assigned to assist a patient recuperating from a head injury with multiple trauma make the transition from the hospital to live with her family. Although the patient, Sara, is physically ready to leave, she suffers from memory impairment, confusion, and slurred speech. Sara is angry about being assigned a case manager, maintaining that she can take care of herself and does not need anyone to help.
INSTRUCTIONS ⇨	Think about what you know about facilitating services for service participants and working with people with disabilities. Identify the necessary individuals and organizations that should be involved in helping Sara to successfully move into her new home and lifestyle. Describe the steps you would take to ensure all parties work together as a team in developing and implementing an individual plan.
EVALUATION CRITERIA ⇨	To receive a proficient rating on this task, you must show all of the following: 1. Knowledge of: • developing and implementing an individualized plan • teamwork and communication skills 2. Ability to propose an effective response to this scenario 3. Ability to communicate clearly in writing

The Process Used to Develop the Written Scenario Prototypes

The process for developing written scenario prototypes involves a number of key steps. They include:

- involve all stakeholders;
- identify and weigh importance of standards;
- develop draft written scenarios;
- conduct reviews;
- conduct pilot test; and
- refine items based on pilot test.

This six-step process is described in detail in **Table 3**.

Table 3.
The Process Used to Develop the Written Scenario Prototypes

Developing Written Scenario Prototypes	
Tasks	**Strategies**
Involve All Stakeholders	• Input from multiple stakeholders ensures that assessments are amenable to a variety of settings, as well as relevant and closely aligned to curriculum and training programs.
Identify and Weigh Importance of Standards	• Identifying and weighting standards is important to (1) target the standards that are most important to key stakeholders and (2) ensure that the selected standards are amenable to written scenario assessments.
Develop Draft Written Scenarios	• Both industry representative and educators should work together to draft scenarios ideas; this ensures that the scenarios will accurately reflect industry contexts as well as match school curriculum.
Conduct Scenario Reviews	• Review of draft scenarios by content experts prior to pilot testing is essential to ensure that all items adhere to effective development guidelines, are clearly linked to targeted standards, and are free of potential bias. • The input received during this period is essential in making sure scenarios are as realistic and relevant as possible. It is important to get a range of reviewers so that the language and prompt is appropriate and familiar to respondents across various administration sites.
Conduct Pilot Test	• Pilot testing is an important part of ensuring the validity of the written scenarios and their accessibility for different respondents. • The "tryout" of items should be done with a sample that is representative of the population for which the assessment is intended. A major purpose of pilot testing is to identify items that do and do not work as intended.
Refine Items Based on Pilot Test	• The information collected during the pilot testing should be analyzed and the results used to improve weak items and inform subsequent development. • If the intent is to develop a full assessment and certification system, then the next steps include continued item development, followed by a large-scale field test.

The experiences of NSSAC using this process to develop and pilot test cross-industry written scenarios is described below.

The NSSAC Experience:

Step 1: Involve all stakeholders

NSSAC made a concerted effort to bring together a diverse group of key stakeholders, including representatives from the four partner industries as well as educators in secondary and post-secondary schools. These industry groups assisted in developing and reviewing scenarios.

Stakeholder input typically included — but was not limited to — determining relative weights of standards for purposes of assessments, developing ideas for items, identifying appropriate scope of practice, ensuring that industry-appropriate language was used, developing scoring criteria, and ensuring that assessments measure the intended outcome (i.e., content and performance standards).

Step 2: Identify and weigh importance of standards

Project partners (American Electronics Association, Human Services Research Institute, National Retail Federation, and WestEd) identified those cross-industry categories that they felt were (1) most important to their industry and (2) most amenable to written scenario assessments. Scenario development then focused on the top-ranked cross-industry categories: *Client Orientation* and *Teamwork*. Identifying and weighting standards can be instrumental in deciding what assessment methods will be employed. Heavily weighted standards might call for more than one type of assessment to capture both the breadth and depth of a respondent's understanding and performance in relation to the standard.

Step 3: Develop draft written scenarios

NSSAC educator and industry groups varied in the extent of their involvement in developing the cross-industry scenarios. For example, human services had education and industry representatives draft extended scenarios which project staff then put into an assessment format. In contrast, retail had project staff draft preliminary generic scenarios and then education and industry representatives added specific industry concepts and ideas.

As in any development effort, more items or prompts were developed than needed for the final assessments. It is not unusual to develop three times the number of prompts you will finally need. Many items may not perform well or as expected in pilot and field testing, thus requiring additional development work or cutting them from the assessment altogether.

Step 4: Conduct scenario reviews

Reviewers for the cross-industry scenarios provided feedback, comments, and suggestions through the use of a review worksheet modified from the scenario development guidelines. This tool allowed reviewers systematically to evaluate each scenario. (See **Appendix B** for a sample scenario review form). For clarification or extended comments, NSSAC staff and project partners conducted follow-up telephone interviews.

Step 5: Conduct pilot test

NSSAC's pilot testing included over 400 students and workers across four industries. Most pilots were education sites, rather than industry sites.

The electronics industry planned to include work sites in piloting, but company upheaval and time demands made participation difficult for some companies. Instead, a key respondent at one site responded to a scenario and gave detailed feedback.

Step 6: Refine items based on pilot test

Many NSSAC revisions involved adding more detail or additional information in describing the scenario situation. For example, respondents found one human services scenario unusually difficult. This particular scenario dealt with arranging services for a client recuperating from a head injury and multiple trauma. Many respondents reported that they were unfamiliar with this type of particular problem and associated service agencies (they usually worked with other types of disabilities). Subsequent revisions by our human services partner helped address such concerns.

Sometimes respondents suggested new and different situations. Other suggestions included clearer directions, additional time for response, giving several scenario options to choose from, making scenarios harder, and providing objective questions instead of or as part of the scenario.

Many respondents thought that no changes should be made. In fact, one industry respondent appreciated how the scenarios demonstrated that even "routine" activities can be complex.

NSSAC's Pilot-Testing Efforts

As outlined above, NSSAC pilot testing included over 400 students and workers across four industries. **Table 4** displays the pilot test sample for each industry by type of site and gender.

Table 4.
Pilot Test Sample for Each Partner Industry by Gender and Type of Pilot Site

Industry Partner	High School or Regional Occupational Center	Post-Secondary	Work Site
Health Care			
Male	*18*	—	—
Female	*84*	—	—
No response	*5*	—	—
Health Care Total	*107*	*0*	*0*
Retail			
Male	*90*	—	—
Female	*145*	—	—
No response	—	—	—
Retail Total	*235*	*0*	*0*
Human Services			
Male	—	*10*	*8*
Female	—	*48*	*10*
No response	—	—	*3*
Human Services Total	*0*	*58*	*21*
Electronics			
Male	*19*	—	*1*
Female	*2*	—	—
No response	—	—	—
Electronics Total	*21*	—	*1*
Grand Total	*363*	*58*	*22*

As shown in **Table 4**, three industries piloted in high schools and regional occupational training centers and one industry, human

services, piloted at post-secondary and work sites. Our pilot test sample was predominately female; 65 percent female versus 35 percent male.

For any one administration, two written scenarios appropriate for that group were alternated so that one-half of the group responded to one and one-half of the group responded to the other. In this manner, we were able to make sure that performance differences on the scenario were related more to the task than to the particular group responding to a certain scenario.

In addition, some NSSAC pilot sites also conducted practice sessions with "rehearsal" scenarios to help familiarize respondents with the test format and expectations for what a proficient response might look like. Teachers and supervisors conducted practice sessions in a variety of ways. For instance, although the written scenario is administered in a set amount of time, practice scenarios can be administered under less stringent guidelines. That is, administrators can read through the scenario prompt and instructions with respondents, answer any questions during the practice administration, and allow extra time to complete the task. Furthermore, administrators may want to give respondents time in groups to discuss their responses to practice scenarios and to offer each other feedback and suggestions for improvements. Several NSSAC pilot sites used the practice scenarios in brainstorming sessions where the entire group developed a solution together. One NSSAC pilot site drew together students from across Louisiana and used the rehearsal as a basis for role-play simulations conducted in small groups led by industry professionals.

A variety of different external factors can effect individuals' performance on assessment tasks. To allow for fair and unbiased evaluations of performance, it is important to know how the assessment was conducted and any special conditions or circumstances regarding the assessment. Thus, it is helpful to ask test administrators to complete a questionnaire describing the assessment situation. This was a strategy we used in pilot testing the NSSAC scenarios.

Most of the feedback from site administrators focused on the timing of administration and respondents' motivation to perform well. Of the administrators responding to the questionnaire, all generally thought that the tasks were appropriate and well-constructed. Some expressed concerns about students' or

workers' unfamiliarity with this particular type of writing task and felt that they might perform better in an oral response format. Other feedback involved potential bias against visually-impaired students/workers when the administrator was not prepared to make alternative arrangements.

During pilot testing it is also important to obtain feedback on tasks from respondents. Thus, we asked students/workers to respond to a few questions about their experience developing and writing a response. We asked (1) whether they had been exposed to the content knowledge required to respond to the scenario, (2) if the task was difficult, and (3) if the task was interesting. Ninety-two percent of the students answered the feedback questions. Overall, respondents were positive about their experiences.

- Sixty-six percent of the respondents offering feedback thought they had been exposed, through training or work, to the occupational knowledge required to respond to the scenario.
- Seventy-two percent of the respondents indicated that the task was not difficult for them. Respondents indicated that the level of perceived difficulty related to (1) experiencing a similar situation before, (2) knowing the content and being prepared, (3) using common sense, (4) having a good teacher, and (5) having experience (or not) with the scenario format.
- Sixty-eight percent of the respondents indicated that they found the task interesting. Reasons given included (1) personal issues, such as liking to work with people or having children the same age as a scenario character, (2) providing an opportunity to practice skills they will need on the job, (3) providing an opportunity to self-evaluate or to test themselves, (4) liking or not liking the context of the scenarios, and (5) providing problems that "made me think" or "let you be creative."

Issues in Developing Cross-Industry Written Scenarios

Development and implementation of the scenarios for cross-industry purposes raised some interesting considerations as related to portability within and across industries.

- *Scope of practice* - Industries, and occupational positions within them, are structured quite differently. Industry skill standards reflect this difference. For instance, the retail skill standards are written for a specific job — professional sales associate — whereas the health care skill standards feature both core and occupational cluster standards. Although the NSSAC health care scenarios were written to the core skill standards (e.g., teamwork), a specific job/position was used to contextualize the scenario prompt. Specifically, pilot-test health care scenarios placed the respondent in the role of a patient care technician, a newer multi-skilled worker designation. There is a trend in the industry in some regions to support and preferentially employ these types of workers. However, those pilot sites in areas where more traditional types of worker designations were the norm had to pre-assess the scenario tasks to decide if their program or training (e.g., certified nursing assistant) related to the same level of performance expectations. For example, although a patient care technician and a certified nursing assistant both perform overlapping tasks, this is not true of all tasks for which each type of worker is responsible. When locally developed or adapted, written scenarios could be used to assess knowledge of job specific skills and competencies.

  Although scope of practice issues were most evident in the health care industry, they were evident in other industries as well. Some retail scenario respondents said they would not be called upon to set up store displays or that their place of employment did not set up displays, such as the one posed in the scenario. So, even though the retail standards are written for a single position, a scope of practice issue emerged from the scenarios written to target this set of standards.

  Currently, nationally validated retail skill standards reflect the needs of an entire industry, and each store, region, or sector of the industry has its own policies that might differ from the national skill standards or from store to store or area to area.

- *Level of generality vs. specificity* - Development efforts focused on providing specific situations yet making them generic enough to cut across the industry. This posed some concerns; some respondents mentioned they needed more information before responding (e.g., human services), while others commented that the topic (e.g., electronics) was too narrow.

This is to be expected when developing assessments for large industries. For example, in human services some community support service workers work with mentally retarded patients while others work in child welfare environments. Developing scenarios that were amenable across these different settings, yet provided enough detail for a realistic context, was challenging.

One possible strategy to address the generality vs. specificity tension is to allow respondents to tailor their scenario responses to the exact context within which they operate. For example, a human services scenario might describe a self-advocacy situation. The respondent might be instructed to select the type of patient he or she works with and then talk about appropriate steps to take in relation to this targeted population. The challenge here, of course, is developing scenarios that allow for these type of options yet still are comparable across employment situations.

The Written Scenario Prototypes

A prototype is a typical representative of a category of object, behavior, or event. Often, it is difficult to find a single example that, by itself, is representative or includes all the characteristics of a particular group. Instead we turn to generic or average prototypes that are a blend of our experiences with many examples in a category. This is the approach we employed in creating cross-industry written scenarios. As we worked with our industry partners to develop written scenarios, we focused on the generic types of work-related problem-solving situations that seemed applicable in more than one industry. This became an informal typology of work-related problem-solving tasks that we applied across industry contexts.

One of the observed challenges facing construction of problem-solving tasks is gauging the difficulty of the task (Woolfolk, 1993). For instance, the presence of relevant versus irrelevant information can affect task difficulty. When a problem is described with irrelevant information included as part of the task, it becomes more difficult to solve. The irrelevant information can become a "red herring" that sets the problem solver off in the wrong direction. Unfortunately, it is sometimes hard to see during written scenario development what constitutes relevant versus irrelevant supporting information.

Other elements to consider include:

- *General versus contextualized problem solving* — Experts in problem solving differ in opinion as to whether general, heuristic processes can be transferred to more specific problem domains. Experts in a domain are more efficient problem solvers because they can define problems quickly and automatically link each problem to a repertoire of appropriate solution strategies. Novices, on the other hand, seem to be best served by relying on a few "tried, but true" general strategies that can be employed regardless of the problem context.
- *Novel versus familiar situations* — Novel problems are more difficult because the solver can't rely on their previous experiences or their schemas built from prior experience until they recognize the problem as similar to one they have solved before.
- *Simple versus complex propositions* — Situations involving multiple relationships are more complex and difficult to solve because the solver must use multiple steps to form a conclusion.

Both problem construction and problem solving can be made easier by holding some elements constant (i.e., only including relevant details, limiting the complexity of the propositions, and using commonplace but not necessarily familiar problems). It can also be helpful to develop prototypical problems which involve different solution processes.

We identified five different prototypes of problem solving tasks which are described in the following section. Not all of these types of problems apply across every industry. Where appropriate, a discussion of responses is included.

Means-End Scenarios

Means-end scenarios require the use of general problem-solving strategies in a skill area related to a specific job. Respondents are given a hypothetical situation in which they are trying to reach some goal and need to find a way for getting there. Instructions for the scenario usually prime the respondent to use means-end analysis to develop a solution by breaking the end goal into subgoals that need to be discussed in the response as well. This type of scenario is the broadest application of our prototypes. All

NSSAC industry partners developed and piloted means-end scenarios.

Figure 17 presents a retail scenario, *Display Dilemma,* which addresses one module of the NRF Skill Standards, setting up and monitoring displays. Specifically, the module is used as the basis for assessing a sales associate's skills in tailoring displays to teenage customers. In writing a response, a respondent must draw on prior knowledge to design a display geared to teenagers, think of ways to judge the display's effectiveness, and plan how different information gained through this process would be used to improve the display. Each of these steps is specified as a point to be addressed in the response. Thus the ultimate goal, an effective display for teenager customers, is broken into sequential steps for the respondent to address.

Different issues may need to be considered at each step. For example, a respondent might say that color, style, and placards or signs should be considered during display design. For the second step, monitoring effectiveness, customer reaction to the display, and daily sales totals could be checked. Finally, checking before and after sales for an item could inform designing future displays.

Figure 17.
Display Dilemma —
A Client Orientation Scenario in Retail

Display Dilemma

Scenario

You are a sales associate in a large department store. Your store is having its annual weekend sale, and you are expecting a large number of teenage customers. As you arrive for work you notice your department's sale display was not constructed the previous night. Your supervisor asks you to spend the morning setting up the display.

Instructions

Think about what you know about presentation and displays. Describe in detail what you need to consider as you set up the display and explain your ideas. Also, describe several ways to monitor the effectiveness of the display during the sale. Discuss the information each monitoring activity would provide and what you could do to improve the display.

To receive a Proficient rating on this task, you must show all of the following:

1. Knowledge of:
   - monitoring effectiveness of displays
   - arranging merchandise to maximize sales
2. Ability to propose an effective solution to this scenario
3. Ability to communicate clearly in writing

Crisis and Follow-up Scenarios

Crisis and follow-up scenarios present a critical situation typical in each industry that needs to be resolved immediately. Instructions in the scenario ask the respondents to tell how they would deal with the crisis and then what long-term strategies they would use to maintain order or prevent reoccurrence. For NSSAC, three out of the four participating industries developed and piloted this type of scenario (electronics, retail, and health care).

Our example of a crisis scenario is *A Change Order*, a teamwork scenario which was developed and piloted with the electronics industry. This scenario, shown in **Figure 18**, presents the respondent with the problem of receiving a manufacturing change on

an important order on a short deadline. The situation involves solving this immediate problem, as well as telling what steps to take to maintain positive team spirit and group cohesiveness. Instructions in this type of scenario ask the respondent to address both the short-term and long-term aspects of the problem in presenting a solution.

Figure 18.
A Change Order —
A Teamwork Scenario in Electronics

A Change Order

Scenario

You are a team leader for a computer hardware manufacturing unit. You have just received a change order on a motherboard that your unit is fabricating. You know this motherboard is part of a large order for a customer and needs to be shipped overseas before 5:00 PM today.

Instructions

Think about what you know about initiating rapid change in the manufacturing process and building team support. Describe the steps you would take to identify the required changes and carry out the new specifications. Also, describe how you would maintain team spirit and cohesiveness during this process.

To receive a Proficient rating on this task, you must show all of the following:

1. Knowledge of:
   - change processes in manufacturing
   - leading work teams
2. Ability to propose an effective solution to this scenario
3. Ability to communicate clearly in writing

' Roles and Responsibilities

This type of scenario focuses on situations in which respondents are requested to delineate the roles and responsibilities of different players in the workplace. The scenarios often involve knowledge of the scope of practice for specific jobs and how that relates to accomplishment of a larger task. Some may relate to line of authority situations, such as the supervisor's responsibility for accurate work and task completion.

The health care scenario shown in **Figure 19**, *The Working Supervisor*, presents a situation in which the respondent's nursing supervisor is busy and unable to attend to patient care. When one of the supervisor's patients calls for immediate attention, the respondent must develop a solution that manages the situation, ensuring that proper procedures are followed and patient needs are met.

In this situation, the respondent may be able to assist by providing some, but not all, of the same functions as the supervisor. Respondents must demonstrate knowledge of appropriate patient care and what they could or could not do, as well as how they might best contribute to the team effort. For instance, a respondent might say he or she would answer the call light and let the supervisor know, make the patient comfortable without administering medication (citing specific activities to accomplish this), and/or find an individual who can administer medication or other therapies, as needed.

These types of problem situations, although common, may not be universal among industries. Scope of practice issues may be more applicable in those industries where a variety of jobs or positions are involved in entry-level work. Only two NSSAC industries, electronics and health care, developed and piloted roles and responsibilities scenarios.

Figure 19.
The Working Supervisor —
A Teamwork Scenario in Health Care

The Working Supervisor

Scenario

You are a patient care technician. Your supervisor, a charge nurse, is assigned patients in addition to her other duties. This shift, she has spent a lot of time on the phone with a doctor dealing with necessary changes to patient's orders. While your supervisor is busy on the phone, one of her patients has turned on the call light. You observe that your supervisor is not responding to the patient's call light.

Instructions

Think about what you know about working as a member of a health care team and the roles and responsibilities of each team member. Given your scope of practice as a PCT, explain how you would ensure that appropriate patient care is given in this situation. Give one example of each of the following kinds of patient care you could administer on your own, could administer with approval, or could **not** administer (someone else would need to do) in this situation.

To receive a Proficient rating on this task, you must show all of the following:

1. Knowledge of:
   - scope of practice
   - teamwork skills
2. Ability to propose an effective solution to this scenario
3. Ability to communicate clearly in writing

Developing Recommendations Scenarios

Developing recommendations scenarios present a situation in which the respondent is required to assess the extent or nature of a problem, draw on his or her knowledge of supports related to the problem, and then compile a list of recommendations for the client/customer to follow. Typically the recommendations involve actions or resources and the respondent is usually asked to provide a reason or justification for each. Two NSSAC industries used this type of scenario, human services and retail. Providing advice based on evaluation of a problem is a universal and, thus, a cross-industry type of situation.

Our example of a recommendations scenario is *Supporting Participant Self-Advocacy*, a human services scenario (see **Figure 20**). This scenario involves a community support service worker who

needs to help a client procure services for bathroom modification. The respondent must develop a solution that includes several recommendations for strategies the client can use to get assistance. In pilot testing, respondents suggested strategies such as getting a copy of the lease or rental agreement, contacting outside agencies, and/or moving to a new apartment.

Figure 20.
Supporting Participant Self-Advocacy —
A Client Orientation Scenario in Human Services

Supporting Participant Self-Advocacy

Scenario

You work as a personal assistance practitioner providing support to people with disabilities who live independently. John, a new participant receiving support services, told you during an initial conversation that the bathroom facilities in his condominium were not suitably adapted to his needs. It was difficult for John to use the bathroom under these conditions. Because his housing arrangement had been classified as "accessible," John wanted to ask the condominium management for additional modifications.

Instructions

Think about what you know about supporting participants with disabilities in obtaining assistance from state, community, and health support agencies. Describe in detail what you should do and say so that John fully understands what options are available for getting assistance and what strategies he might use for obtaining aid to get his bathroom facilities appropriately adapted.

To receive a Proficient rating on this task, you must show all of the following:

1. Knowledge of:
   - self-advocacy methods and support groups
   - how to increase awareness of self-advocacy methods and groups
   - adapting interactions to best match participant needs
2. Ability to propose an effective solution to this scenario
3. Ability to communicate clearly in writing

Competing Clients or Priorities Scenarios

The competing clients or priorities scenario presents the equivalent of an in-basket task for line workers. Developing and presenting a solution involves juggling conflicting priorities along more than one dimension. Respondents have to prioritize competing tasks, keeping their clients in mind, and then explain how

they would communicate their decision to the clients. Two of the NSSAC industries developed and piloted these type of scenarios, health care and electronics. We expect, however, that prioritizing tasks is fairly universal among industries.

The competing clients or priorities scenario in **Figure 21**, *Prioritizing Patients' Needs*, presents the situation of a health care technician who is responsible for several patients with different kinds of needs and different levels of criticality. A newly arrived patient requires that specimens for lab tests be collected that will require personal attention for at least a half hour. Meanwhile, another patient is angry and demanding attention. The respondent must determine which patient and tasks need to be attended to first and then communicate his or her decision to the patients.

Many respondents develop solutions that involve settling the angry patient down by explaining why they were delayed, bringing him juice or a snack, and then explaining to him that another patient needs their attention. Respondents state that they plan to return, after collecting the required lab samples, to perform their regular duties and other types of patient care.

Figure 21.
Prioritizing Patients' Needs —
A Client Orientation Scenario in Health Care

Prioritizing Patients' Needs

Scenario

You are a patient care technician in a busy Intensive Care Unit (ICU) for an acute care facility. Today, you have four patients. This includes a new patient, just admitted to the unit. The new patient, Ms. Turrell, needs to have urine and blood specimens collected as soon as possible. Mr. Jones, another patient, has had his call light on for the last five minutes. So far you have been unable to answer it. Mr. Jones is now angrily yelling that he wants his dinner.

Instructions

Think about what you know about communication and patient care. Describe in detail what you should do so that patient care activities are prioritized in an appropriate manner. Explain what you would do or say to make sure Mr. Jones's needs are recognized.

To receive a Proficient rating on this task, you must show all of the following:

1. Knowledge of client interaction including:
   - providing personal care for patients
   - communicating effectively
2. Ability to propose an effective solution to this scenario
3. Ability to communicate clearly in writing

Using the Written Scenario Prototypes for Assessment Development

Although a variety of prototypes may be developed for cross-industry written scenarios, we found five generalized models that are applicable across two or more industries. Our prototypes include: means-end, crisis and follow-up, roles and responsibilities, developing recommendations, and competing clients or priorities problem situations. These situations provide a range of cognitively complex tasks to challenge respondents to apply general as well as work-related knowledge and skills. Additional prototypes may be included as different industries begin to focus on written scenario development. The cross-industry written scenario prototypes can be used as templates for constructing industry-specific contexts for such scenarios.

Portfolios

Unlike scenarios, *portfolios* are a cumulative assessment, representing student or employee performance over time. Portfolios are valued for their flexibility and their ability to demonstrate

accomplished learning in virtually any industry or academic discipline. By containing a range of both written and nonwritten products, portfolios can effectively assess achievement and competence in relation to desired standards.

A "showcase" portfolio, sometimes included as part of certification or high stakes assessment systems, is a *purposeful* collection of evidence. It is designed to show students' or workers' ability to apply their knowledge and skills by developing "real life" work samples in their occupational area. Such a portfolio contains polished work or finished products that incorporate feedback and revisions rather than daily work samples.

Within an assessment and certification system, portfolios can follow students or workers from year to year as they accumulate evidence of their mastery of occupation-related skills and competencies. If the portfolio is comprehensive, it can show evidence of achievement in a range of areas involving both career preparation and the academic disciplines. Students and workers can adapt such a comprehensive collection and develop different "showcase" portfolios for different potential employers.

Educational Purposes for Portfolios

In education and school-to-work settings, portfolios provide a framework for organizing and presenting the work that takes place both in and outside of the classroom. Teachers can examine what they are already doing with students to identify work assignments that will naturally generate portfolio entries. Similarly, students with jobs or internships can identify work products that may be appropriate for their portfolios. In this way, the work that goes into the portfolio can be an outgrowth of normal classroom learning activities and practicum placements, not separate tasks or tests created just for the purpose of assessment (WestEd, 1998).

In general, portfolios in school-to-work and community college settings reflect the following characteristics. They:

- ***show work-in-progress***
  Portfolios can inform teaching and learning processes, as well as provide cumulative evidence documenting achievement of certain standards. Instructional use can be enhanced by conferencing strategies between teachers and students and periodic review of student work.

- ***incorporate work-related experiences***
  This might include job apprenticeships, community service experiences, mentorships, and internships. Over one-third of the resources listed in a recently released toolkit for assessment of work-based learning are portfolio-based (Northwest Regional Educational Laboratory, 1996).

- ***include a self-assessment component***
  Portfolios involve students in a variety of higher level cognitive processes, or metacognition. Students engage in processes such as planning, goal-setting, allocating time and resources, negotiating conflicting priorities, and analyzing the strengths and weaknesses of their own work.

  By reflecting on their work, students can more effectively think about what they accomplished in relation to standards, decide how more effective work might be accomplished, and plan further work or learning activities. Students also reflect on how they might best present themselves to employers as they try to enter or move within the workforce.

- ***serve as a vehicle for communication***
  Portfolios can serve as a vehicle to stimulate interaction and conversations between faculty, schools, and employers. Often school-to-work or occupational training programs are not fully connected to regular faculty and instruction. Others working at the school site may not value the quality of education in such programs or the depth of knowledge and skills needed for some tasks and jobs. Such marginalization can leave these programs vulnerable to budget cuts and reduce teachers' opportunities to collaborate and to reinforce skills and concepts across disciplines and learning contexts (Grubb & Badway, 1995).

Education portfolios geared for school-to-work preparation may focus on workplace readiness or career preparation while still incorporating academic skills in many ways. For example, the Career-Technical Assessment Program (C–TAP) portfolio, one of the components of the overall C-TAP assessment system, includes a range of items designed to support education, specifically as it relates to career-preparation (Rabinowitz, 1997). **Figure 22** lists the C–TAP portfolio components.

Figure 22.
Components of the C-TAP Portfolio

The C-TAP portfolio includes:

- a ***letter of introduction***, including a description of the contents of the portfolio and career aspirations, as well as an analysis of the work included in the portfolio
- a ***career development package***, consisting of a resume, job or college application, and a letter of recommendation
- ***work samples*** linked to key program content standards, demonstrating hands-on ability to apply knowledge and skills to job-related tasks
- a ***writing sample*** on a topic that is integral to the students' targeted career path
- a ***Supervised Practical Experience evaluation form*** related to a work-based learning experience. A supervisor rates the students' performance in terms of employability skills such as interpersonal skills and occupational safety

Using the C-TAP portfolio as a model, one NSSAC pilot site pilot tested a portfolio for use in health career preparation at an occupational training center. The portfolio, adapted specifically for dental assistant students, provided evidence of a range of occupational skills learned in the classroom, including the ability to use technology, and included industry-specific certificates. The required contents for this portfolio can be seen in **Figure 23**.

BEST COPY AVAILABLE

Figure 23.
Components of the Student Dental Assistant Portfolio[2]

The Dental Assistant Portfolio includes:

Certificates

- computer skills, registered dental assistant skills, dental radiography, cardiopulmonary resuscitation card

Occupational competencies

- completed checklist for classroom / lab skills
- clinical observation samples including 1) evaluation from workplace supervisor and 2) student written report from clinical observation
- work sample demonstrating computer skills (e.g., database)

Job search skills

- completed employment application
- resume (also demonstrating word-processing skills)
- letter of introduction (not included in sample)

The occupational competencies section of the dental assistant portfolio is similar to the work sample components of the C-TAP portfolio and can provide documentation of hands-on job related skills and competencies. This feature is readily adaptable across different industry areas. For example, as shown in **Figure 24** and **Figure 25**, a clinical observation work sample documents accomplishment of specific tasks relevant to dental assistant competencies through supervisor observation sign-off (**Figure 24**) and a written student report (**Figure 25**). In this example, the student has observed only some of the necessary tasks. Depending on the nature of the program and level of student capacity and background, expectations for those aspects of performance a student might address in the written report may vary.

[2] Courtesy of Pam Cooper, Lead Teacher, Santee Health Occupational Center.

Figure 24.
Community Classroom Dental Assistant
Dental Assistant Student Evaluation[3]

COMMUNITY CLASSROOM
DENTAL ASSISTANT
DENTAL ASSISTANT STUDENT EVALUATION

Student: *Sheila Thomas* Date Initiated: *4-3-96*

Assistant's Name: _____

Please grade assistant trainee by checking appropriate square. A fair and sincere evaluation will be appreciated and will help the trainee become an efficient dental assistant.

	Good	Average	Poor	Not Observed
1. Personal appearance	✔			
2. Ability to handle patients				✔
3. Ability to follow directions	✔			
4. Attitude	*Great!*			
5. Chairside techniques				✔
A. Aspirating				✔
B. Use of syringe				✔
C. Maintain doctor's visibility				✔
D. Anticipation of doctor's needs				✔
E. Transfer of instruments				✔
6. Tray set-up technique	✔			
A. Knowledge of procedures	✔			
B. Instrument identification	✔			

The student completed the above skills with 80% accuracy. _____
Supervisor's Signature

Sheila is obviously very interested in what we are doing and asks very appropriate, intelligent questions. It's a pleasure having her observe.

[3] Courtesy of Pam Cooper, Lead Teacher, Santee Health Occupational Center.

Figure 25.
Dental Student Written Report from Clinical Observation[4]

I observed in oral surgery and passed instruments. After we scrubbed-up and set-up the area I learned all the instruments by number. That was quite interesting after finally learning them by their names. Everyone was ***extremely*** *concerned about cross-contamination. Your scrubs can't even touch the table your instruments are on.*

I enjoyed oral surgery. The pace was nice because you set-up and then talked with the patient and explained what would be taking place. Cmdr. Young was fun to work with. The whole group was pretty silly.

For retail, a similar type of work sample might relate to sales presentation skills. For instance, one NSSAC retail teacher uses a classroom activity that could easily fit into the work sample portfolio component. Pairs of students are required to role-play an entire sales presentation consisting of four parts, Approach, Desire, Objections and Close. The simulations are videotaped and the teacher rates their performance using the following guidelines (see **Figure 26**).

[4]Courtesy of Pam Cooper, Lead Teacher, Santee Health Occupational Center.

Figure 26.
Sales Presentation Rating Form [5]

Salesperson _____

Date _____

Approach

1. Appropriate dress (5)
2. Offer firm handshake, present business cared, and ask to be seated (5)
3. Establish rapport (5)
4. State purpose (3)
5. General benefit statement (5)
6. Reasons for needing information (1)
7. Current situation questions (5)

Securing Desire

8. Familiarization (5)
9. Ask questions to identify needs/problems (10)
   (Minimum of two questions per need)
10. Confirm prospect's desire to solve problems (5)
11. Provide features, benefits, proof (15)
12. Use of visual aids (5)
13. Trial close after each benefit (5)

Handling Objections

14. Acknowledge prospect's concern after objection (5)
15. Clarify each objection (5)
16. Provide features, benefits and proof (15)
17. Use of visual aids (5)
18. Offer trial close after each objection (5)

Closing

19. Proper use of closing technique (5)
20. Wait for response, recommend an order or a plan of action, and confirm the details
    (If an order is taking place, did the sales person have another form?
    If the product is going to be resold, were profits per unit and total profits mentioned?) (10)
21. Reassure prospect and state follow-up plan, thank prospect, and exit properly (5)

Other Skills

22. Maintain prospect's interest and attention during presentation and get prospect involved (4)
23. Smile (2)
24. Use of prospect's name (3)
25. Objectionable mannerisms (5)
26. Enthusiasm (2)
27. Confidence and poise (2)
28. Eye contact (3)

Total Points *(possible are in parentheses)* (150)

Excessive Timeouts _____

Net Points _____

[5] Courtesy of Larry Wharton, Director, Business Leadership and Management Program, Mt. Hood Community College.

Industry Purposes for Portfolios

Portfolios are not new to industry. Architects, photographers, and artists have produced portfolios to showcase their talents and competencies for some time. A free-lance artist, for example, must have samples of prior work on hand to show prospective clients or funders. Although the actual contents of a portfolio varies according to profession and specialty, each person typically includes a variety of items showing that he or she has delivered consistently high quality work over a period of time.

In the case of a free-lance artist, a portfolio of work will be judged against the criteria a specific client or funder has for a specific job. Industry portfolios also may be judged against criteria unique to individual clients. However, portfolios often are evaluated based on criteria developed by a training program or company. For example, Beth Israel Hospital has developed guidelines for assembling portfolios for nurses seeking advancement through their Professional Nurse Advancement and Recognition Program. Specific criteria exist for advancement to higher levels of responsibility. See **Figure 27** for requirements for the Clinical Nurse IV level portfolio.

Figure 27.
Requirements for the Clinical Nurse IV Portfolio at Beth Israel Hospital

Required components include:

- two clinical exemplars (narrative accounts of practice that answer questions about the significance of and decision making involved in a critical event)
- a performance evaluation from a supervisor
- a self-evaluation
- two letters of reference from nurses with a specified level of experience (one currently employed at the hospital)
- an application for advancement stating professional goals

Optional components include:

- a curriculum vita
- examples of other related work and training.

BEST COPY AVAILABLE

In NSSAC, an interesting variation linking work experience (e.g., internships) to school programs emerged in one of the pilot sites from the human services industry. At this site, community college teachers and industry representatives developed a curriculum based on practicum experiences with families in a Head Start Program. The curriculum was based directly on the Human Services Community Support Provider skill standards. Course assignments addressed specific competency areas. A student portfolio used for evaluation captured ongoing documentation for work experiences (e.g., log, journal, schedules), professional development, training, and experience documentation (copies of transcripts, professional development plan, personal goals for self-growth), and the teacher's evaluation summary of the portfolio. Overall, one-half of all possible points for course grading were based on the portfolio. Portfolio requirements for this course are detailed below (see **Figure 28**).

Figure 28.
Greater Lawrence Headstart-Human Service Practicum Portfolio Requirements

Section A:

- Time Sheets/Document Hours
- Time Management Log Assessment/Work
- Time Management-Work Activity list
- Written Weekly Logs as Assigned
- Log Contains Content Required

Section B:

- Personal Philosophy Human Services
- Copy of the Headstart Mission Statement
- Resume
- Copies of Transcripts/Courses
- Staff Development Plan/Prof Development Plan
- Headstart Personnel Performance Evaluation
- Personal Goals for Self Growth
- Copy of Job Description/Supervision Record
- Staff Survey

Section C:

- Evaluation of Portfolio Assignment Based on Skill Standards

TOTAL POINTS FOR GRADE _____

FACULTY _____

Each of these portfolio examples have implications for a cross-industry assessment and certification program. For instance, our example industry portfolios include some evidence of direct observation of performance by an independent assessor or supervisor. This is a reasonable expectation for an industry context, given that many workers do all or parts of their work in isolation. Verification of accuracy or precision is an important quality control function. As the workforce becomes increasingly professionalized, this aspect of job performance can be expected to become more reliant on workers themselves and, thus, external accountability is increasingly important at specific points along a career pathway.

A Cross-Industry Portfolio Template

Creating a portfolio template for use in employment and educational settings poses many challenges. Identifying parameters for portfolio entries that meet the needs of all stakeholders requires dialogue and continued feedback. Ideally, a cross-industry portfolio template should be broad enough in scope so that it is adaptable to the various settings. In addition, the portfolio entries should represent general categories of work to help ensure portability across industry and educational settings.

Developing the Cross-Industry Portfolio Template

Given the diversity of settings and purposes for assessment in industry alone, NSSAC staff endeavored early on in the development process to engage industry representatives in discussion about assessment practices in their specific industry. Two methods for obtaining feedback from industry representatives were used: *expert panel groups* and *employer surveys*. In both cases, we attempted to elicit specific feedback about the feasibility of using a portfolio in their industry. This feedback, coupled with related work on the Career-Technical Assessment Program (C-TAP), served as the basis for developing a cross-industry portfolio template.

Expert Panel Groups

NSSAC project staff and partners convened four separate expert panel groups — one per industry — consisting primarily of industry representatives. Expert panel members reviewed a sample student portfolio based on the C-TAP model (see **Figure 22** above), commented on the portfolio entries, and offered suggestions as to how (if possible) to make them more relevant to industry. A total of 33 industry representatives and educators participated in the expert panels; most of the industry representatives worked in human resources and/or education and training positions.

The expert panel group members varied in their comments about the portfolios. Although portfolios were generally viewed as positive and valuable for students, their use in industry was not always seen as practical. In fact, one participant suggested that having to look through so much paper and documentation might even create a negative bias for job applicants. Positive aspects related to students showing initiative and work effort by "putting something like that together" and thinking through their

own accomplishments and what that might mean to potential employers (see **Table 5** for a summary of the expert panel groups and employer survey).

Employer Survey

An employer survey was distributed to a sample of employers in each of the four partner industries. A total of 43 employers responded to the survey. Respondents represented companies that ranged in size between 17 to 200,000 employees, although three-fourths of them had 1,000 or fewer employees. Of the 43 respondents, 21 had participated in the expert panel groups, while the additional 22 respondents had been surveyed independently by the American Electronics Association (see **Appendix B** for survey results).

The survey asked respondents to indicate (1) how they valued different information sources and (2) how they used each source of information (i.e., resume, work samples). Results from the survey are consistent with the findings from the expert panel review. For example, interviews, personal contact with employee references, resumes, and work samples were highly valued by both the expert panel groups and survey respondents.

As compared to other personnel-related uses, respondents reported that the information sources were most frequently used for hiring purposes. The most commonly used information sources for advancement were interviews, resumes, and performance tasks. Finally, the most commonly used information sources for post-training assessment included interviews, resumes, performance tasks and writing samples.

Table 5.
Summary of Industry Perspectives on Portfolio Components
(from expert panel groups and survey)

Information Source	Summary of Employer Perspectives
Resumes	The *resume* was highly valued. Participants from the electronics industry expressed concern that *resumes* reflect up-to-date skill sets, such as expertise with software applications, programming code, network architecture, website development, and e-mail addresses.
Cover letters	Participants did not find cover letters or letters of introduction to the portfolio particularly valuable.
Letters of references	Expert panel group participants did not find *letters of recommendation* particularly valuable as they were duplicative with the resume. Survey respondents placed only slightly more value in them.
Phone contact with personal references	Employers valued *personal references* from other employers. They especially appreciated having enough information so that they could contact the references, ask questions, and get employers' and clients' feedback. For the health care industry, references indicated the quality of a program or links to an established network of co-workers.
Interviews	*Interviews* were universally used and valued. Often, job applicants were asked to role-play or solve hypothetical problem situations on-the-spot. A few participants from the retail industry commonly put applicants on the floor for a brief period to show customer service skills when they come in for an interview.
Performance tasks (e.g., simulations, work samples)	The *work samples* were highly valued although participants suggested ways to better tailor the work samples for industry use. Some participants liked that teachers signed-off on the work samples, ensuring they were original work.
Computer-based assessments, writing samples, and licenses and certifications	Employers somewhat valued these information sources but did not rely on them alone.
Other	Two industries, human services and retail, relied on customer feedback, either through customer response forms or extended probationary periods. A small number of employers listed additional information sources including driving records, criminal record searches, and drug tests.

The Cross-Industry Portfolio Template

The results from the focus groups and surveys informed the development of the cross-industry portfolio template which is streamlined to facilitate industry use. Most of the participating employers contended that education-based portfolios too often contain information sources (e.g., cover letters, letters of recommendation, and writing samples) that are not highly valued by or feasible to industry.

Employers may be more likely to use portfolios that feature their more valued information sources, such as resumes, performance tasks, and interviews. The cross-industry portfolio template below (see **Figure 29**) attempts to address the needs of both industry representatives and educators.

Figure 29.
A Cross-Industry Portfolio Template

Career Development Package

- *Resume* - This component should profile skills and experiences related to the job the student/worker is seeking.
- *References Contact List* - Names and contact information for two professional and one personal reference with a brief description of the context of their relationship with the student/worker for each.

Checklist of Skill and Competencies

- For employees, the company skills list can be used, as informed by the relevant national skill standards, to develop a critical competencies list. Students/workers include a brief description of what they did that demonstrated their mastery of a competency with teacher or supervisor sign off on each competency.

Industry Scenarios

- One or two completed scenario assessments which may be written or involve another format for the student's/worker's response. Potential employers may choose to use these scenarios as a basis for asking applicants to orally respond to new or different problems during an interview.

Work Samples

- Two to three samples of work related to either the student/worker's current job or anticipated one. One of these work samples may involve the use of technology. Each work sample should be accompanied by an abbreviated version of a summary sheet. The summary should describe the work sample and how it relates to the skill standards for the industry, and incorporate a teacher/supervisor sign-off.

This template is easily tailored to different industries and potential jobs, as well as education sites. The components reflect information sources identified by the industry experts as valuable and used for a variety of decision-making purposes.

Although NSSAC pursued pilot test efforts with portfolios over a limited span of time, preliminary results from other related efforts also suggest that employers across industries hold similar views about what types of information should inform hiring, promotional, and training decisions.

BEST COPY AVAILABLE

Chapter 6: Scoring and Reporting

"As an educator, I found the entire process interesting, especially the scoring session which gave me a new perspective about how to interpret student responses.

Often, students are not prepared to take tests. My students benefited from the preparation they received from the rehearsal scenarios. It was a very positive element that is all too regularly lacking from the students' educational experience."

— Joe Cotey
READY Teacher
Harper High School
Chicago, Illinois

Developing an effective scoring system is a complex and challenging process, one that often gets insufficient attention during assessment development. Yet, assigning scores and reporting the results to key stakeholders are both integral parts of any assessment. Scoring systems are a means of interpreting the relationship between standards and individual achievement. Depending on the nature of the assessment being scored, the scoring system can be very simple and straightforward or very complex. Scoring assessments such as multiple-choice tests is simple; the scorer only has to determine whether or not the respondent selected the pre-determined correct response. This straightforward scoring process is usually done by machines. On the other hand, scoring performance-based assessments, such as written scenarios or portfolios, involves evaluating an individual's open-ended response. In this situation, there is no one right answer, rather there may be several different ways of producing correct answers. Thus, these assessments require that the scorer interpret a person's response relative to predetermined criteria.

This chapter focuses on scoring of assessments, such as written scenarios and portfolios, that require an individual to independently produce a response. It draws heavily from previous work done by WestEd for the California Career-Technical Assessment Program (WestEd, 1998) and includes examples specific to the cross-industry NSSAC effort. We begin with a brief overview of developing an effective scoring system, followed by a discussion on issues related to developing cross-industry assessments, as informed by the work of NSSAC. The chapter concludes with a discussion on the reporting of scoring results. While the scoring principles laid out in this chapter are intended to provide a general overview for those interested in designing assessment systems, the complex, technical scoring-related procedures and analyses required of high stakes assessment are beyond the scope of this document.

Developing a Scoring System

The development of a scoring system goes hand-in-hand with the development of assessment tasks. As such, it is an iterative process: items or tasks are drafted, tried out, analyzed based on their scoring results, revised, and tried out again until the standards, the assessment instrument, and the scoring system achieve a satisfactory match. Within this process, there are specific activities aimed at ensuring an effective scoring system, whether in an educational or employment setting. These include:

- developing a scoring plan;
- drafting scoring scales for performance assessments;
- checking for validity;
- checking for reliability; and
- choosing a cut score to reflect the performance standard.

Developing a Scoring Plan

The development of a scoring system begins with identifying the standards to be addressed by the assessment. A scoring plan identifies both the specific standards-based knowledge and skills to be measured by the assessment as well as the method(s) of scoring to be used.

For written scenarios, the scoring plan can include the number or percentages of items that address particular standards. For portfolios, the scoring plan might identify how the targeted standards or scoring dimensions are reflected in the portfolio entries that comprise the assessment. **Table 6** shows an example of a scoring "map" that bridges a scoring plan with the overall assessment strategy.

Table 6.
Map of Dimensions to Portfolio Entries

	APPLICATION	LETTER OF RECOMMENDATION	RESUME	WORK SAMPLES	WRITING SAMPLES	SUPERVISED PRACTICAL EXPERIENCE EVALUATION	TABLE OF CONTENTS	LETTER OF INTRODUCTION
CONTENT • Knowledge of major ideas and concepts in career-technical standards • Knowledge of how skills in career-technical areas are applied				X	X	(X)		
CAREER PREPARATION • Career planning • Personal qualities needed for employment	X	X	X	(X)		X		X
ANALYSIS • Evaluation of own skills and work • Investigation and information gathering			X	X	X	(X)		X
COMMUNICATION • Attention to audience • Using own ideas • Organization and clarity • Accuracy, neatness, and completeness • Language mechanics, sentence vocabulary	X		X	X	X		X	X

93

Source: WestEd, 1998

The table indicates which entries in the portfolio are likely to provide certain information. As shown, assessment dimensions (topics reflective of individual or combinations of standards) are listed on the left side of the map, and the headings at the top refer to the different entries of the C-TAP portfolio assessment. Each X indicates a portfolio entry that is designed to produce responses related to a specific scoring dimension.

The scoring plan should also specify the method of scoring to be used. Two major types of methods are *holistic* and *analytic* scoring. *Holistic* scoring views an individual response to an assessment as a whole or as an integrated performance. A scorer considers information about specific aspects of the performance only as it contributes to the overall impression left by the entire performance. The holistic scoring method results in a single score or narrative description that summarizes the performance as a whole.

In contrast, the *analytic* scoring method views a student's response to an assessment in parts. A scorer rates different aspects of a performance separately and usually combines these separate ratings into an overall score. Sometimes, certain aspects of performance are deemed to be more important than others and therefore their scores are given more weight when calculating the overall score. (Examples of each type of scoring method will be presented and discussed in greater detail in the next section.)

Drafting Scoring Scales for Performance Assessments

The next step is to draft a scoring scale for each assessment used. A scoring scale is a system of classifying assessment performances in a progressive series of points, levels, or degrees. Scoring scales, which are basically descriptors of different levels of student performance, may be developed for use with the holistic or analytic scoring method. For example, **Figure 30** illustrates the holistic scoring scale for the C-TAP portfolio assessment as a whole.

Figure 30.
C-TAP Portfolio Holistic Scoring Scale

ADVANCED: Shows superior content knowledge and application of knowledge and skills related to the career-technical standard(s); shows superior ability to prepare for a career; demonstrates superior self-evaluation skills; overall presentation is well-organized and effective, making all of the work easy to understand.

PROFICIENT: Shows adequate content knowledge and application of knowledge and skills related to the career-technical standard(s); shows adequate ability to prepare for a career; demonstrates adequate self-evaluation skills; overall presentation is organized, making most of the work easy to understand.

BASIC: Shows gaps in content knowledge and/or application of content knowledge and skills related to the career-technical standard(s); shows some ability to prepare for a career, but major weakness(es) may be evident; demonstrates vague or sketchy self-evaluation skills; overall presentation makes some of the work difficult to understand.

LIMITED: Shows little or no content knowledge and application of content knowledge and skills related to the career-technical standard(s); shows little or no ability to prepare for a career; self-evaluation skills are weak; fails to present work effectively.

The four points in the scoring scale include two levels of satisfactory performance: *Advanced*, where the individual has met the standards with distinction; and *Proficient*, where the individual has met the standards but not excelled. The other two score points represent levels of unsatisfactory performance: *Basic*, where the individual does not meet the standards at the present time, but shows promise of meeting the standards with some additional focused work; and *Limited*, where the individual does not come close to meeting the standards and may need substantial remediation. Note that these performance levels are ordered such that each level of the scoring scale represents a point on a continuum ranging from weaker to stronger performances.

Table 7 presents an example of an analytic scoring guide for different aspects of a writing assessment.

BEST COPY AVAILABLE

Table 7.
Sample Analytic Scoring Guide for a Writing Assessment

IDEA DEVELOPMENT

1 Little or no elaboration of ideas

2 Simple, general or repetitious elaboration of ideas

3 Appropriate and detailed elaboration of ideas

ORGANIZATION

1 Minimal or no discernible structure

2 Clear, simple, or repetitious structure

3 Adequate and suitable structure

FACILITY WITH LANGUAGE

1 Extremely simple style, with little or no voice

2 A more varied style, with identifiable personal voice

3 Interesting and effective style, leading to a clear personal voice

4 Precise, controlled, and pleasing style, leading to a strong personal voice

MECHANICS

1 Mechanics make reading difficult

2 Mechanics seldom interfere with reading

3 Mechanics are appropriate and do not interfere with reading

4 Mechanics are well-controlled and enhance the writer's ideas

NOT RATABLE

These writing samples are unratable because they are blank, completely illegible, written in a language other than English or entirely disregard the writing prompt.

Taken from Welch et al., 1994.
as cited in WestEd, 1998

Each aspect uses a different scale, ranging from 1-3 to 1-4 points. Using this scoring method, the separate aspects of performance (idea development, organization, facility with language, and mechanics) receive separate scores that can be combined for a total score.

Whether opting for the holistic or analytic scoring method, assessment developers can often draft the scoring scales at the same time as assessment tasks. If the developer is not sufficiently familiar with the range of responses likely to be produced by students or workers, scoring scales should be developed during the item tryout, or pilot test, phase instead. Either way, several examples of student or worker responses should be collected which clearly represent strong, average, and weak performances. These examples should then be analyzed against the relevant content standards to identify the specific characteristics that distinguish the different performance levels.

In fact, to help communicate their meaning, scoring scales for performance-based assessments are usually accompanied by examples of responses, called benchmarks, that illustrate each level of performance. The benchmarks provide concrete examples of the scoring scale descriptors and offer an opportunity for stakeholders to understand more fully the meaning of each score or score level.

Labeling of performance levels on a scale is also very important. Typically two types of scales are used: (1) descriptive labels, such as *Basic, Proficient,* or *Advanced;* or (2) a numerical scale, such as "0-4" or "1-6" linked to performance levels. Both types of scales represent attempts to make meaningful qualitative distinctions among responses.

Developers also should be aware that the first time a performance-based assessment is given, scores may be depressed because of respondents' lack of familiarity both with the standards being assessed as well as the assessment format. Over time, however, as trainers, educators, and examinees become more familiar with the standards upon which the assessments are based and as examinees become more familiar with the assessment format, both instruction and responses typically improve. For this reason, it is important to leave some room at the highest levels of the scoring scale for improvement during later years of implementation.

Finally, in drafting scoring scales it is sometimes unclear how many performance levels are optimal for feedback and reporting. However, there are indicators when the number of performance levels is inappropriate. Each score should produce a meaningful distinction between performances. If scorers are having trouble distinguishing between adjacent scores (even after repeated refinement to the scoring scale) it is often a sign that the number of scoring points should be reduced. Conversely, if scorers are feeling uncomfortable with the range of performances represented by a single score, it is a sign that the number should be increased.

Checking for Validity

Scoring provides accurate information about students or workers only if the assessment instruments and their scoring scales accurately reflect the particular standard(s) being assessed. When such alignment is present the scoring system is said to be valid. Checking for alignment, or validity, helps ensure that the assessment tasks will elicit performances that provide scorable evidence relevant to specific standards.

One method of checking the validity of scoring scales is review by content experts (e.g., workers, supervisors, educators) of assessment tasks in relation to their scoring scales. This involves having a committee of content experts make a judgment as to whether the assessment tasks and scoring scales accurately reflect the standards targeted for assessment. Another method for investigating validity of the scoring system involves statistical review of assessment response data. Although the full details of this complex process are beyond the scope of this introductory document, there are several key factors to consider when reviewing response data for validity. For example, it is important to investigate whether or not the assessment tasks are at the appropriate difficulty level. This involves checking the distribution of scores on the different levels of a scoring scale. The distribution of scores for tasks should also be disaggregated by such variables as gender and ethnicity to identify group differences in performance. Tasks for which there are large performance differences should be reviewed by content experts to see if the content of the task is biased against a particular group.

Checking for Reliability

In addition to validity, assessment developers should check for the reliability of an assessment scoring system. As previously indicated, reliability is the degree of confidence that both scores and performances are replicable over time and across different circumstances. Replicability of scores means that the same response will receive the same score(s) no matter who scores it or when it is scored. For example, different scorers or the same scorer at a different time should assign the same scores to a given response. Replicability of performances also means that individuals will perform similarly on different tasks designed to measure the same standard at the same level of difficulty.

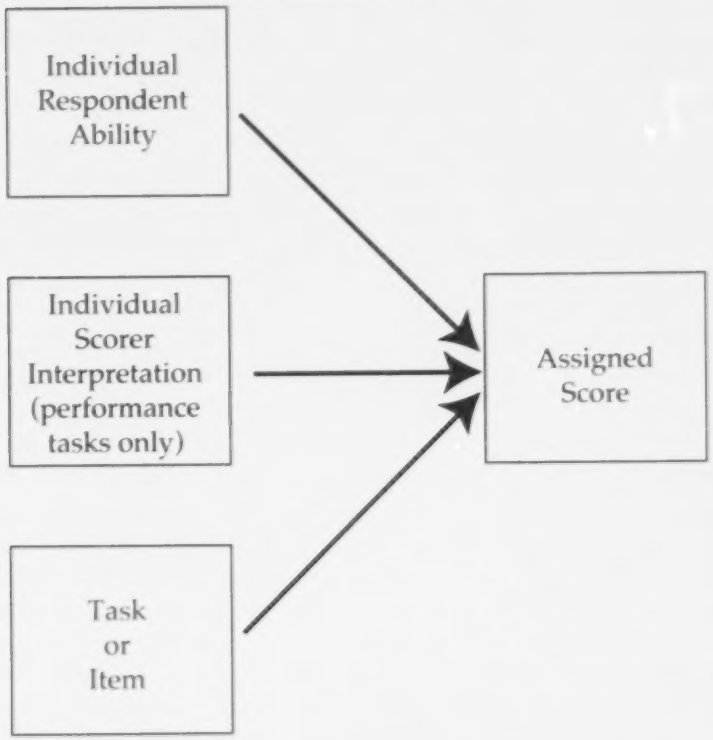
Achieving reliability for performance-based assessments is a challenge, because scoring such assessments requires interpretation and professional judgment. Once scoring scales are developed, their meaning must be accurately communicated to the scorers, so that scorers can both understand and internalize them. In large-scale, high-stakes performance tasks, training sessions provide scorers with structured opportunities to become familiar with the scoring scales and their application to the particular tasks being scored. Understanding of the scoring scale is usually further facilitated by the systematic review of sample responses designated as *benchmarks* which are accompanied by written and/or oral explanations of how each sample response reflects the relevant level on the scoring scale. As mentioned previously, the benchmark responses represent different points on the scoring scale. In training, scorers are taught how to apply the scoring scale to the benchmark responses.

During the scoring of high-stakes, performance-based assessments, several reliability checks are built into the process. For example, samples of responses are scored by more than one scorer (i.e., double scored) to ensure that the degree of agreement among scorers is high enough to meet standards of reliability generally accepted in the field. In addition, individual scorers are often checked for *drift* from a correct understanding of the scoring scale by having them score previously scored responses.

Large-scale assessments also check for *generalizability* of scores, or how replicable performance is across similar tasks. Generalizability studies statistically compare the effects of different factors (e.g., the respondent, the scorer, the task) on scores.

This relationship between relevant factors and scores is portrayed in **Figure 31**.

Figure 31.
Generalizability Model



Ideally, an individual respondent's abilities should affect the scores more than a task or a scorer. If individual scorers have a large impact upon scores, then it means that some scorers are consistently scoring more strictly than others. If the task has a large impact, then the performance on that task is unique and does not reliably predict performances on other tasks designed to measure the same set of knowledge and skills. Performance-based assessment tasks tend to have relatively high task effects (i.e., performance varies considerably as a function of the specific task). This is why performance tasks often are used in combination with multiple-choice items for high-stakes assessment systems.

Choosing a Cut Score to Reflect the Performance Standard

The final step to developing an effective scoring system is the selection of a cut score to reflect the performance standard, or expected performance relative to the content standards. The cut score identifies a dividing line where individuals scoring at or above the line are considered to have demonstrated mastery relative to the content standards while those scoring below the line have not. For performance-based assessments, a single point or level of the scoring scale is chosen to reflect satisfactory mastery of the relevant standard(s). For example, C-TAP assessments set the cut score at the *Proficient* level. Individuals scoring at the *Proficient* level and above on the scoring scale are demonstrating mastery of the content standards, while individuals scoring at the *Limited* level and below are not. Cut scores also are often set as points other than mastery to signal exemplary performance or notable progress towards mastery.

The choice of the cut score exemplifying the performance standard is typically based on the judgment of content experts and informed by empirical (field test or actual test) results. The task of the content experts is to reach a consensus on the specific score that best represents satisfactory performance relative to one or more standards.

For an assessment system consisting of various assessment methods and tasks, expected results across the different assessment components are combined to yield an overall cut score. Again, content expert judgment and empirical test results are often used to determine the relative weights of the various assessment components and in determining the final composite cut score.

Scoring Written Scenario Assessments — the NSSAC Example

NSSAC activities included a formal scoring process for retail written scenarios. The highlights of this process are presented below, providing a concrete example of the scoring steps and principles described above. (Please note that the final step, *Choosing a cut score,* is omitted because this step typically applies to an instrument consisting of several assessment items and tasks as opposed to an individual assessment task. The NSSAC effort focused on pilot testing of individual tasks as opposed to an overall instrument.)

Step 1: Develop the NSSAC written scenario scoring plan

As alluded to in a previous chapter, the decision was made early in the project to focus written scenario assessment development on the cross-industry standards areas of *Client Orientation* and *Teamwork*. Such standards areas were deemed by content and assessment experts involved with the project to be amenable to written scenario assessment. The decision also was made to use a holistic scoring method modeled after C-TAP written scenario scoring scales because of the empirically demonstrated validity and reliability of C-TAP scoring processes and scales.

Step 2: Draft scoring scales for NSSAC written scenarios

Figure 32 shows the general written scenario scoring scale used which specifies four levels of performance, ranging from *Limited* to *Advanced*.

Figure 32.
Sample of a Holistic Scoring Scale

C-TAP Generic Written Scenario Scoring Scale
Part II: Overall (Holistic) Score

Score Levels			
Limited	**Basic**	**Proficient**	**Advanced**
•Shows little or no knowledge of required career-technical standards*; may or may not address scenario requirements*; supporting arguments (if required) are poor or lacking; writing problems interfere with communication of ideas	•Shows gaps in knowledge of required career-technical standards*; may or may not address scenario requirements'; supporting arguments (if required) are sketchy or vague; writing problems sometimes interfere with communication of ideas	•Shows knowledge of all required career-technical standards*; addresses all scenario requirements adequately or most requirements fully*; supporting arguments (if required) are logical and presented with adequate detail; communicates ideas effectively	•Shows superior knowledge of all required career-technical standards*; fully addresses all scenario requirements*; supporting arguments (if required) are well-developed; communicates ideas effectively

* Career-technical standards and scenario requirements depend on the specific scenario.

Based on your examination of the overall student response and the dimensional scores for the response, please review the overall (holistic) scoring levels described above and give the response a final, overall score. Please provide a justification for the overall (holistic) score.

Overall Score:_ _____

Justification: _____

_____

_____

_____

Student ID Number: _____

School/Work Site: _____ Evaluator: _____

While the general scale affords the opportunity to use the same criteria for scoring regardless of the scenario prompt, NSSAC also developed item-specific scoring scales. Item-specific scoring scales use the general performance level definitions as a basis for creating such definitions for a particular prompt. **Figures 33** and **34** show an example of a retail scenario prompt and its item-specific scoring scale, respectively. Consistent with the procedures described above, this prompt and its scoring scale were refined in a recursive process with educators and industry representatives.

Figure 33.
Display Dilemma Scenario Prompt and Instructions

Display Dilemma

Scenario

You are a sales associate in a large department store. Your store is having its annual weekend sale, and you are expecting a large number of teenage customers. As you arrive for work you notice your department's sale display was not constructed the previous night. Your supervisor asks you to spend the morning setting up the display.

Instructions

Think about what you know about presentation and displays. Describe in detail what you need to consider as you set up the display and explain your ideas. Also, describe several ways to monitor the effectiveness of the display during the sale. Discuss the information each monitoring activity would provide and what you could do to improve the display.

To receive a Proficient rating on this task, you must show all of the following:

1. Knowledge of:
   - monitoring effectiveness of displays
   - arranging merchandise to maximize sales
2. Ability to propose an effective solution to this scenario
3. Ability to communicate clearly in writing

Figure 34.
Item-Specific Scoring Scale

	Display Dilemma
4	• Considers the target market when describing 3 or more factors for planning the display with some explanation. • Describes 2 or more ways to monitor the effectiveness of the display. • Offers 1 or more ideas to improve the display AND 1 or more types of information.
3	• Considers the target market when describing 2 or more factors for planning the display and may or may not explain ideas. • Describes 1 or more ways to monitor the effectiveness of the display. • Offers 1 or more ideas to improve the display OR 1 or more types of information.
2	• Considers the target market when describing 1 factor OR describes 2 or more factors without tying to the target market. • Describes 1 or more ways to monitor the display OR offers 1 or more ideas to improve the display OR offers 1 or more types of information.
1	• Describes 1 factor which may or may not be tied to the target market. OR • Describes 1 or more ways to monitor the display OR offers 1 or more ideas to improve the display OR 1 or more types of information.
0	Incorrect or irrelevant response

Figure 34. (continued)
Item-Specific Scoring Scale

Examples of factors:

- Plan
- Colors
- Style
- Price of sales items
- Shape of display
- Location
- Complimentary items
- Signs to include
- Props
- Teen fashions and trends
- Tools to assemble
- Season of the year

Ways to monitor:	Information provided:
• Watch customers	• See if looking/are attracted
• Watch customers (window display)	• See if more are entering store
• Question customers	• What they think, like, might suggest
• Given surveys to customers	• What they think, like, might suggest
• Compare sales	• See if sales increased with display

Examples of ideas to improve an effective display:

- Add/delete items based on need
- Change the props (or the design of the items)
- Place signs in front of the store or on the street to attract more customers
- Move the display to a different part of the store
- Restock, reorder

Step 3: Check for validity

To help ensure the validity of the scoring scales, the NSSAC written scenario prompts and item-specific scoring scales were subjected to review by content experts at various points in the development process, including initial drafting of items, revision prior to pilot-testing, and benchmarking/scoring.

Step 4: Check for reliability

For NSSAC, a formal scoring session was conducted focusing on two retail written scenarios — *Understaffed* and *Display Dilemma*. The scoring session was conducted in Chicago over a three-day period, with representation from each retail pilot site. Six content experts participated in the scoring process: one retailer, four high school retail teachers and a retail program evaluator. The first two days were spent benchmarking the scenarios (selecting sample responses that reflect the various levels of the scoring scale and are used to train scorers). The final day was spent training the scorers and scoring student responses to the two scenarios.

The goal of the benchmarking session was to review and select responses for each item to be scored. These responses were then used to introduce and train scorers on the following day. As an introduction to the benchmarking process, participants were provided with a review of the items, the general four-point scoring scale, and the item-specific scoring scales for each item. The benchmarking process included selection of the following types of responses:

Anchor papers — solid examples of performance at each rating level.

Training papers — responses that illustrate scoring issues and challenges.

Calibrating papers — good examples of performance at each rating level. These papers are used to test that scorers understand the item-specific scoring scale and scoring process.

Figure 35 shows an example of an anchor paper for score level "4" (Advanced) for the *Display Dilemma* scenario. The justification for this score, showing how the response reflects important standards-based knowledge and anaylsis skills (Client Orientation, Arranging Merchandise, and Monitoring Effectiveness), is also presented.

Figure 35.
Example of Score Level "4" Response and Justification

"Display Dilemma" Written Scenario
Student Response

Before I would start to set up my display, I would first figure out what items I would need. After finding the items that I would need, I would make a display plan. The first step in my plan would be to choose a open or closed display depending on the price and prestige of my items. The next step would be to choose a display style. Abstract would probably be my choose due to the fact that mostly teenagers will be at the sale. Another important step would be to choose vibrant eye catching colors to attract attention. Also signs that can be clearly and quickly read should be placed, to let the customer know your item is on sale. Since the visual attention time a customer gives a display is between 6 to 8 seconds the color and signs are important factors. After the plan has been made and all these factors have been addressed its time to create your display.

After your display is complete it is important to maintain it. The display should maintain a clean and neat order at all times. After customers look through your display it is important to straighten it up so it will be neat for the next customer.

Your display's effectiveness can be measured many ways. One way to know if your display is effective is to watch to see if customers are stoping to look at your display. Another way to measure it success is by the products sold. But if your display is not effective the best solution would be to revise your display style or the colors.

Setting up a display can be difficult. But with the right knowledge and good planning it can be a easy success.

Content Knowledge and Analysis

- **Client Orientation**: Identifies the target market — teenage customers — and selects display elements appropriate for this group.
- **Arranging Merchandise**: Demonstrates general knowledge about important factors to consider when arranging a display, particularly in relation to teenage customers.
- **Monitoring Effectiveness**: Discusses two ways to monitor the effectiveness of the display. Also, offers an improvement strategy for the display if it is not effective.

Following the benchmarking process, a scoring session was held in order to score the written scenario responses. A *table leader* was selected to facilitate scoring for each of the two retail written scenarios. The table leaders, both of whom had participated in the benchmarking process, were responsible for training the scorers and also explaining the rationales for the benchmark selections.

Prior to scoring the written scenarios, participants were asked to do the following:

- review and discuss the item-specific scale;
- review the anchor papers as a group and discuss;
- score the training papers one at a time using the item-specific scale and discuss as a group;
- score the first set of calibrating papers; and
- score the second set of calibrating papers.

The goal is for potential scorers to correctly score at a minimum of 80% agreement on the calibration papers to demonstrate their understanding of the scoring process. For both scenarios, scorers attained 80% agreement between scorers on the last set of calibration papers.

In all, scorers rated a total of 235 responses. **Table 8** shows the distribution of scores for each scenario. Overall, 16% of the students who responded to *Display Dilemma* and 12% of the students who responded to *Understaffed* scored at the Proficient level or above (rated as a "3" or "4"). The fairly low percentages of Proficient and Advanced scores for both scenarios are consistent with the previously mentioned finding that performance on new types of assessment modes tend to be somewhat depressed when they are first introduced.

Table 8.
Distribution of Scores for the Two Retail Scenarios (n = 235)

	Display Dilemma (n = 113)		**Understaffed (n = 122)**	
Score Level	*n*	%	*n*	%
0 (Irrelevant)	12	11%	4	3%
1 (Limited)	61	54%	64	52%
2 (Basic)	22	19%	40	33%
3 (Proficient)	16	14%	7	6%
4 (Advanced)	2	2%	7	6%

Reporting Assessment Results

After an assessment is scored, the next major step is to communicate the results to the appropriate audience(s). Assessment information can be reported in a variety of ways. A common format is *numeric scores* that summarize an individual's performance. Types of numeric scores frequently used to report results include raw scores, percentage scores, percentiles, or scale scores.

A *raw score* is how many assessment items a respondent answered correctly. It is only informative if it is reported along with the total number of items in the assessment. A *percentage score* indicates the percentage of items that a respondent answered correctly, thus providing more information than a raw score. For example, reporting a score of 30 percent conveys more information than reporting a raw score of 30. Another type of numeric score, *percentiles*, report a respondent's performance relative to others. For example, a respondent scoring at the 80th percentile did as well as or better than 80 percent of the sample who participated in norming the test. Finally, *scaled scores* are similar to percentiles in that they also take into account the mean and standard distribution of the population norming the test. However, scaled scores can use predetermined ranges of scores that vary widely. For example, the Scholastic Assessment Test uses a scale ranging from 200 to 800.

In contrast to numeric reporting formats, *developmental continua*, often in the form of checklists, can offer specific information about performance. **Table 9** provides an example of such a checklist, describing specific knowledge, skills, and abilities to be mastered by the respondent.

Table 9.
Example of a Developmental Continuum (Checklist) for Work Ethic Standards

	Exceeds Standards	Meets Standards	Below Standards
Makes decisions quickly after due time is given to fact-finding and consideration of the alternatives.	_____	_____	_____
When necessary, disagrees and debates with others in a professional, respectful manner and always uses positive methods of persuasion.	_____	_____	_____

From Academy High School Internship Preparation Program, cited in Bailey and McTighe, 1996, p. 123.

In summary, it is critical that the assessment reports and the process that produces them should be made meaningful to important stakeholders, such as employers, workers, students, parents, teachers, and the general public. Most stakeholders may not take the time to achieve the depth of understanding needed to fully interpret assessment reports. However, if a reporting system is meaningful and sound, the pieces will be in place to satisfy those with questions or concerns and to ensure that the information is accurate and able to contribute to appropriate decisions.

Chapter 7: Lessons Learned and Implications for Designing a Comprehensive Skill Standards-Based Assessment System

"The NSSAC WestEd Scenario was a beneficial assessment tool for the READY students. The rehearsal and written scenarios allowed students to develop, analyze and demonstrate their written communication skills in relation to their retail experiences. This WestEd tool prepared students for solving realistic problems encountered in the world of work."

— Barbara Donaldson
READY Program
Director

Although this document is intended primarily as an implementation guide, it also represents the culmination of a research project on skill standards-based assessment. As such, it is appropriate to end with a summary of "lessons learned" from NSSAC and specific recommendations for design and implementation of a skill standards-based assessment system.

Lessons Learned

Lesson 1: While there is notable overlap among skill standards across industries, many stakeholders want assessments that are tailored to the specific context of their industry.

Although NSSAC focused on four sets of industry skill standards (electronics, health care, human services, retail) that differed substantially in structure, language, and level of specificity, comparative analysis yielded substantial commonalities across these sets of standards. Specifically, the project identified eight cross-industry competency areas (*Client Orientation, Teamwork, Evaluating and Interpreting Information*, etc.). Several similar studies have also found, through inductive or empirical methods, much content overlap across sets of industry skill standards.

A major implication of this finding for cross-industry assessment is the possibility of using the same tasks or instruments to assess workers' competencies in different industries. Using the same instruments for different industries would result in development and administration cost savings. More importantly, it would mean that workers assessed for skills in one industry might receive "credit" for skill attainment in another industry.

The current study did *not* lend support to the possibility of occupational assessment tools with portability across industries. Project participants (project partners, assessment task developers, pilot test participants, employer focus group participants) early on in the collaborative effort expressed their strong preference for couching assessment tasks in the context of their particular industry. The reasons given for this preference related to both *content validity* and *face validity* considerations. A major content validity consideration is that knowledge and skills are inextricably bound to context. For example, participants felt that although an emphasis on teamwork skills is common across industries, the specific teamwork skills differ across industries. Similarly, a face validity consideration is that in order for an assessment task to be credible with the stakeholders in a particular industry, that

assessment tool must look like it is assessing skills specific to that industry. In fact, even within the same industry or occupational cluster it often proved difficult to develop assessment tasks (scenarios) that were deemed by content experts to have content and face validity for more than one specific occupational area.

Lesson 2: Employers and educators differ in meaningful ways with respect to their perspectives on assessment.

Just as industry-specific concerns must be addressed in any cross-industry assessment effort, substantial differences in assessment perspectives amongst employers and educators also must be considered. A clear lesson learned from the NSSAC effort is that educators and employers have different views on and needs for assessment. Thus, while there seems to be some consensus among both educators and industry representatives that assessments should be linked to challenging standards, there is no consensus on what constitutes appropriately challenging standards. Among educators, the primary focus is on setting high standards for academic content areas. While vocational educators are beginning to embrace industry skill standards, a number of "mainstream" educators have expressed concern to project staff that most sets of industry skill standards (with the possible exception of those for high tech industries) may inadequately reflect high-level academic knowledge and skills. This is in contrast to employers who generally felt that industry skill standards should be direct, concise, and reflective of workplace needs rather than academic goals.

Another major difference noted between educators and employers who took part in this project is that employers seemed more concerned with legal defensibility issues (specifically, demonstrated job relevance) than educators. This observed difference most probably stems from the fact that the participating educators focused heavily on classroom assessment of student achievement with respect to academic and skill standards, whereas employers expressed concern about worker certification. Clearly, the latter is more vulnerable to legal challenges.

Lesson 3: Given a choice, employers want streamlined assessment procedures and tools.

Another major lesson, closely related to Lesson 2, is that employers want assessments that are easy to administer and concise. For example, employers generally do *not* want to wade through voluminous portfolios of prospective employees to seek evidence

of skill mastery. This viewpoint contrasts with educators who seem to embrace use of portfolios as a means for students to take ownership of and showcase their work.

What are the implications of this lesson for employers' use of portfolios and other more complex assessment tools? First, portfolios might be streamlined to include a combination of checklists of skills, other assessment results, and very targeted work samples. Furthermore, use of portfolios in industry might be limited to specific applications, avoiding uses which involve review of huge numbers of portfolios at one time. For example, it may not be feasible to use portfolios as an initial screening tool for applicants to entry level positions (e.g., sales associates). However, when the applicant pool of tens or hundreds is narrowed down to the most qualified five or ten finalists, it may be prudent for employers to consider portfolios or collections of work to make final selections among a handful of candidates. The additional hands-on information provided by portfolios may help employers make more informed employee selection decisions. Similarly, portfolios may be helpful in decision-making about promotions.

A final consideration is that industries may differ in the applicability of portfolios to employee training, recruitment, selection, or promotion. For example, while it may be unwieldy and inefficient for portfolios to be used in initial screening of retail associates, it may be advantageous to use portfolios to screen potential human services workers. Compared to retail associates, fewer candidates tend to apply for any given human services position, making it more feasible to review portfolios in the application process. Moreover, human services employers may be more motivated to take the time to look at portfolios of applicants since human services workers must be able to deal successfully with sensitive issues in their clients' lives, and a hiring mistake would therefore be very costly.

Lesson 4: There is strong interest in written scenarios for assessment of skills in schools and the workplace.

Unlike portfolios, there was widespread support among both educators and employers for written scenarios. Stakeholders recognize that written scenarios require respondents to apply occupationally-related knowledge to solve realistic problems. Such items, when carefully constructed, can elicit a range of complex cognitive behaviors, for example, organizing,

summarizing, classifying, comparing, evaluating, predicting, concluding, and creating. Moreover, both employers and educators like the written scenario because they represent a relatively cost-efficient means of measuring competency in relation to targeted standards. Finally, the scenario approach affords great flexibility; prompts and responses can be adapted to assessment modes other than written, such as video or oral assessments.

Lesson 5: Common processes and templates can help facilitate the development of assessment tasks across industries and educational settings.

A practical contribution of NSSAC to the national skill standards efforts is the identification of common processes and design of templates that can be used to facilitate the development of standards-based assessment tasks. Chapter 5 presented step-by-step development processes and templates for portfolios and written scenarios that can help expedite the complexities of assessment development. Moreover, common processes and templates can help assure some standardization of assessment tasks across industry and educational applications. By offering common processes and templates to stakeholders, we can help "demystify" assessment for the field and help build recognition and support for skill standards and related assessments.

Conditions for Successful Implementation of Skill Standards-Based Assessment Systems

In addition to the specific lessons learned through this collaborative effort, conclusions were drawn about the kinds of significant changes that need to occur to support implementation of skill standards-based assessment systems. These conditions include: structural changes, technical support, as well as incentives and mandates.

Structural Changes

Specific changes to the infrastructure of industry and education are needed to support a skill standards-based assessment system. These include establishing meaningful education-industry partnerships and professional development opportunities for teachers and employers. Industry-education partnerships must include participation of management as well as practitioners. For example, policy and decision-makers at the highest levels from both sectors are needed to secure the commitment necessary to

develop and support comprehensive assessment systems. The industry-education partnership should also encompass practitioners, providing a forum for industry representatives and teachers to collaborate on the selection or development of assessment tasks.

All active participants in industry-education partnerships will need professional development to orient them to skill standards and assessment. In particular, those participating in assessment development would need substantial professional development to prepare them for design and administration of assessment tasks, as well as skills to ensure alignment of classroom/workplace activities with the assessment tasks.

Technical Support

Technical issues often threaten full implementation of any new or innovative assessment system. As previously discussed, low reliability and generalizability are commonly cited as technical flaws in new, predominantly performance-based assessment systems. Likewise, combining scores across varied assessment tasks to make a certification decision and then equating evolving assessments across years and industries pose challenges to the technical quality of a comprehensive skill standards-based assessment system. Although these are real technical challenges, structural changes (such as opportunity for training and professional development) can help mitigate their potentially negative impact. For example, several studies have demonstrated that the reliability of performance-based assessments increases to an acceptable levels through adequate training of scorers on well-defined scales (Baker, 1992; Shavelson, Baxter, and Gao, 1993). Teachers and workers need time and training to become proficient in scoring and other aspects of assessment development.

The availability of outside assessment expertise also is essential to ensure technical adequacy (Khattri, Reeve, Kane, and Adamson, 1996). Although many companies have the financial resources to obtain outside assessment expertise, few make the necessary investment. Schools, on the other hand, need a variety of different resources to support alternative assessment development and use, including money, expertise, time, and effort.

Incentives and Mandates

Both the development and implementation of a skill standards-based assessment system represent huge undertakings. As such, there need to be strong incentives, including explicit mandates, to support this change. With respect to industry, many companies (larger ones in particular) have their own prehire assessment and evaluation tools and systems in place. These employers may be complacently satisfied and reluctant to try new tools or learn new ways of thinking about performance. Thus, employers must be presented with the value added of adopting new instruments. One incentive is the potential portability of skill standards-based assessment tools. Employers need to be convinced that the new skill standards-based assessment tools are adaptable to already existing company training programs and assessment needs and that the portability afforded by such tools are worth the effort of their adoption. Another incentive for employers is the value added of involving their staff in the development or selection of skill standards-based assessments. For example, high performance workplaces are moving towards peer performance review (American Electronics Association, 1997). In these contexts, employers may be motivated to provide experiences for their employees that would help them to think about workplace performance and assessing their own and others' performance against specified standards.

The endorsement and adoption of skill standards-based assessments by industry may be incentive enough for educators to embrace such systems. Furthermore, local and state mandates can also facilitate the process. For example, the previously described Certificates of Initial and Advanced Mastery (CIM/CAM) were first called for by the Oregon Educational Act for the 21st Century (1991). This act, or mandate, provides a needed catalyst for development of a standards-based, school-to-work assessment system in that state. Finally, formal adoption by a state or locality of national skill standards for one or more industry could also serve as an effective incentive or mandate for the development of a comprehensive skill standards-based assessment system.

Concluding Remarks

Clearly, much work needs to be done to pave the way for the development of comprehensive skill standards-based assessment systems. Although still very much in its infancy stage, the

groundwork is being laid to develop such systems of portable certificates. The NSSAC effort is one example of a research and development effort focused on laying this important groundwork. The National Skill Standards Board (NSSB) is providing support to many other assessment-focused efforts as well, including national voluntary partnerships that are developing comprehensive national skill standards and certification systems. As these national efforts begin to make strides alongside individual state and local efforts, it becomes increasingly clear that significant resources are needed to make the structural changes, obtain technical support, and establish clear incentives and mandates to enable successful implementation of comprehensive skill standards-based assessment systems. We need strong industry-education partnerships coupled with widespread public support to make this happen.

Appendix A The National Skill Standards Board (NSSB)

The National Skill Standards Board (NSSB)

Established by the National Skill Standards Act in 1994, the NSSB is charged with creating a voluntary, national system of skill standards for workers in entry level through first-line supervisory positions. These skill standards are seen as crucial to ensuring a well-trained and competitive American workforce for the global economy. Consisting of 28 representatives from industry, education, and government, the NSSB was formed to oversee and support at the national level the development of a skill standards certification system.

The NSSB does not set the standards for industry or education. Instead, they have:

- identified 15 broad economic sectors in which skill standards will be developed. These areas are closely aligned to traditional industry categories (See **Figure 25** showing the economic sectors).
- promoted the establishment of "voluntary partnerships" — which by law must include employer, union, worker, community, and education and training representatives — to develop skill standards for each of the economic sectors;
- researched and disseminated information on standards development to the voluntary partnerships;
- developed a national framework to support skill standards (see below for more detailed information about this national framework).

Once the skill standards are developed, the NSSB has the responsibility to endorse skill standards created by these voluntary partnerships. Certification systems will be based on the skill standards endorsed for each sector and will provide information about 1) student/worker progress and competence and 2) the effectiveness of occupational training and school-to-work programs (National Skill Standards Board, 1996).

Figure 25.
The NSSB Economic Sectors

Agriculture, Forestry, and Fishing
Business and Administrative Services
Construction
Education and Training
Finance and Insurance
Health and Human Services
Manufacturing, Installation and Repair
Mining
Public Administration, Legal and Protective Services
Restaurants, Lodging, Hospitality and Tourism, and Amusement and Recreation
Retail Trade, Wholesale Trade, Real Estate and Personal Services
Scientific and Technical Services
Telecommunications, Computers, Arts and Entertainment, and Information
Transportation
Utilities and Environmental and Waste Management

The NSSB Framework for Industry Skill Standards

NSSB has developed a common framework for skill standards development, drawing on the insights gained through analysis of state and national skill standards projects. This framework establishes the guidelines NSSB will use to endorse standards for workers in entry-level through first-line supervisory positions within each of the economic sectors.

The NSSB framework for setting standards covers three types of knowledge and skill, ranging from the broad to the specific: *core*, *concentrations*, and *specialties*. Standards for each of these three levels will be described in terms of the *academic* skills and knowledge, the *occupational* skills and knowledge, and the *employability* skills and knowledge required to carry out *critical functions*. "Critical functions" are defined as "the major chunks of work that must be performed and which, taken together, constitute the

critical or principal responsibilities of the individuals involved" (The National Skill Standards Board, 1996).

While the NSSB defined the *core* area for each of the economic sectors, the voluntary partnerships will be responsible for defining the *concentration* and *specialty* levels of the standards. A voluntary partnership might designate between 0 and 6 concentrations, but there will be no limit to the number of specialties.

In order to qualify for NSSB endorsement, the standards created by the voluntary partnerships must:

- follow a common nomenclature identified by NSSB;
- describe in clear terms the critical work functions specific to the core, concentrations, and specialties;
- describe the academic, employability, and occupational knowledge and skills necessary to perform the critical work functions for the core, concentrations and specialties;
- adhere to statutory requirements and Board policy on assessment;
- be consistent with civil rights law;
- meet or exceed the highest applicable standards used in the United States, including registered apprenticeship standards;
- be benchmarked to the best international standards;
- be forward looking; and
- include a plan for the updating and continuous improvement of standards and certificates.

These principles will ensure relevancy, broad-based support, and the ultimate effectiveness of the skill standards movement.

Appendix B
NSSAC Scenario Review Form

National Skill Standards and Assessment Collaborative

Scenario Feedback Form
(Please replicate to give feedback on more than one scenario)

Fax/mail to Mike Crepeau by
March 25, 1997
415-241-2702
Thanks

Your Name _____

I am a(n):

_____ Educator _____ Employer
_____ Practitioner _____ Student
_____ Other (please specify) _____

Title of Scenario: _____

1. Is the problem posed a realistic dilemma requiring applied knowledge in a particular situation or circumstance?

_____ Yes _____ No
Comments:. _____
_____
_____

2. Do the scenario and instructions present a clear, cohesive task to the student?

_____ Yes _____ No
Comments:. _____
_____
_____

3. Are the skills and knowledge required critical to the occupational area?

_____ Yes _____ No
Comments:. _____
_____
_____

4. Do the instruction bullet points highlight the major career performance concepts required in this response?

_____ Yes _____ No
Comments:. _____
_____
_____

5. Is the vocabulary level appropriate for a high school student?

_____ Yes _____ No

Comments:. _____

_____

_____

6. Is the vocabulary level appropriate for a post-secondary student or recent hire?

_____ Yes _____ No

Comments:. _____

_____

_____

7. Will a respondent be able to complete the task in 45 minutes?

_____ Yes _____ No

Comments:. _____

_____

_____

8. Are any diagrams, graphics, articles, resources required to complete the task easily interpreted (i.e., won't take much time to use)?

_____ Yes _____ No

Comments:. _____

_____

_____

9. Are the scenario and instructions of manageable length to be read aloud?

_____ Yes _____ No

Comments:. _____

_____

_____

10. Is the context for the scenario a realistic setting that all students/workers can imagine?

_____ Yes _____ No

Comments:. _____

_____

_____

11. What would you expect to see in a proficient student response to this scenario?

Comments:. _____

_____

_____

Appendix C
Employer Survey Results

Employers' Ratings of the Value of Different Information Sources About Employees' Skill Levels (n = 43)

	Not Valuable		Somewhat Valuable		Very Valuable		Average Rating	
Information Source	n	%	n	%	n	%	x	SD
Job application forms	5	12	9	21	25	58	2.5	.72
Resumes	0	0	10	23	30	70	2.7	.44
Cover letters	13	30	16	37	8	19	1.9	.75
Letters of reference	7	16	16	37	9	21	2.1	.72
Phone contact with personal references	1	2	13	30	25	58	2.6	.54
Interviews	0	0	0	0	39	91	3.0	.00
Performance tasks (e.g. simulations, work sample)	2	5	4	9	22	51	2.7	.60
Writing samples	5	12	10	23	12	28	2.2	.76
Computer-based assessment	4	9	9	21	13	30	2.3	.74
Standardized tests (e.g. multiple choice)	9	21	4	9	5	12	1.8	.88
School transcripts	5	12	12	28	4	9	1.9	.67
Psychological tests	11	26	5	12	4	9	1.6	.81
Licenses/certifications	3	7	13	30	9	21	2.2	.66
Portfolios	5	12	11	26	4	9	1.9	.69
Other	1	2	0	0	3	7	2.5	1.00

Employers' Use of Different Information Sources About Employees' Skill Levels (n = 43)

	Use for hiring	Use for advancement	Use for post-training
Information Source	*n*	*n*	*n*
Job application forms	37	5	2
Resumes	41	16	6
Cover letters	30	0	0
Letters of reference	31	1	1
Phone contact with personal references	39	4	0
Interviews	41	21	8
Performance tasks (e.g. simulations, work sample)	19	12	6
Writing samples	15	9	6
Computer-based assessment	13	5	4
Standardized tests (e.g. multiple choice)	8	3	5
School transcripts	11	1	2
Psychological tests	9	5	2
Licenses/certifications	22	9	4
Portfolios	11	1	0
Other	3	1	0

NSSAC Project Partners

Kathy Mannes
National Retail Federation
Washington, DC

Marianne Taylor
Human Services Research Institute
Cambridge, MA

Cheryl Fields Tyler
Andrea Fiorelli
American Electronics Association
Santa Clara, CA

Policy Advisory Committee

Beverly Campbell
California Department of Education
Sacramento, CA

Lisa Ehrlich
Advanced Systems
Dover, NH

Joan Knapp
Knapp & Associates International, Inc.
Princeton, NJ

Owen Layton
Western Maricopa Consortium
Litchfield Park, AZ

Harry O'Neil
University of Southern California/CRESST
School of Education
Los Angeles, CA

Stanley Rabinowitz
WestEd
San Francisco, CA

Bernie Ronan
Estrella Mt. Community College
Litchfield Park, AZ

Joan Wills
Institute for Educational Leadership
Washington, DC

Health Care

Pam Cooper
Grossmont Health Occupations Center
Santee, CA

Marla Keeth
Pasadena Health Academy
Pasadena, CA

Margery Simchak
Washington High School
Phoenix, AZ

Pat Schneider
Phoenix College
Phoenix, AZ

Cap Tiwald
Centennial High School
Peoria, AZ

Ailey Runyon
Peoria High School
Peoria, AZ

Electronics

Jack Carter
Nevada Union High School
Nevada City, CA

Bob Hoskins
Turlock High School
Turlock, CA

James Callen
Intermec
Everett, WA

Human Services

Mary DiGiovanni
North Essex Community College
Reading, MA

Donna Gray
Middlesex Community College
Lowell, MA

Patty Lauziere
Arc of Greater Lowell
Lowell, MA

Bob Fera
Middlesex Community College
Bedford, MA

Barry Schwartz
Vinfen Corporation
Cambridge, MA

Melissa Fleming
North Essex Community College
Reading, MA

Maggie LaBella
North Shore Community College
Danvers, MA

Lisa Griffith Johnson
Abilities Unlimited—PLUS Program
Westfield, MA

Retail

Carol Borsky
Pauline Perez
Louisiana Business Partnership
Baton Rouge, LA

Clare O'Donnell
READY Program
Chicago, IL 60606

Bibliography

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

American Electronics Association. (1997). *A user's guide for skill standards: Helping educators and trainers compare and evaluate industry skill standards*. Washington, D.C.: NSSB.

Ananda, S. M., Rabinowitz, S. N. (1995). Developing a comprehensive occupational certification system. In Institute for Educational Leadership, *Skill standards: Making the connections in the education and training network.*

Ananda, S. M, & Rabinowitz, S. N. (1996). *Developing a comprehensive occupational certification system*. San Francisco, CA: Far West Laboratory.

Ananda, S. M. (1997). *The ultimate school-to-work challenge: Linking assessment in school and the workplace*. Paper in progress.

Baker, E. L. (1992). *The role of domain specifications in improving the technical quality of performance assessment*. (Tech. Rep.) Los Angeles, CA: UCLA Center for Research on Evaluation Standards and Student Testing.

Bailey, T. R., & Bernhardt, A. D. (1996). *In search of the high road in a low-wage industry*. Institute on Education and the Economy Brief, Number 14. New York: Institute for Education and the Economy.

Beth Israel Hospital. (1993). *Guidelines for Clinical Nurse IV Portfolios*.

Center for Occupational Research and Development (CORD). (1995). *National voluntary skill standards: Hazardous materials management technology*. Washington, D.C.: U. S. Department of Education.

Center for Workforce Development (1996). *In search of commonalities*. Washington, D.C.: Institute for Educational Leadership.

Commission on the Skills of the American Workforce. (1990). *America's choice: High skills or low wages*! Rochester, NY: National Center on Education and the Economy.

Electronic Industries Foundation. (1996). National Skill Standards Development Project. *Study of the state of the art of certification and accreditation programs.*

Far West Laboratory / WestEd. (1995). *National health care skill standards.* San Francisco, CA: FWL.

Grubb, W. N. (1996). *Working in the middle: Strengthening education and training for the mid-skilled labor force.* San Francisco, CA: Jossey-Bass Publishers.

Grubb, W. N. and Badway, N. (1995). *Linking school-based and work-based learning: The implications of La Guardia's co-op seminars for school-to-work programs.* Berkeley, CA: National Center for Research in Vocational Education.

Guttman, Robert M. (1997). *Skill standards and equal opportunity.* Washington, D.C.: U. S. Department of Education.

Holmes, S. (1985). *Test equating issues in certification and licensure testing.* ERIC Document 260 116.

Hoppe, P. and Wills, J. (1996). *Developing a common framework for the skill standards system.* Washington, D.C.: Institute for Educational Leadership.

Human Services Research Institute. (1996). *Community support skill standards: Tools for managing change and achieving outcomes: Skill standards for direct service workers in human services.* Cambridge, MA: Human Services Research Institute.

Institute for Educational Leadership. (1996). *Synthesis of conversations with skill standards pilot projects.* Washington, D.C.: Institute for Educational Leadership.

Institute for Educational Leadership and DynCorp•Meridian. (1996). *Developing Skill Standards: A Technical Manual.* Washington, D.C.: Institute for Educational Leadership.

Jenkins, D. et. al. (1994). *The role of standards, assessment and credentialing in educating the high performance worker: Lessons from Denmark and Great Britain.* Center for Learning and Competitiveness.

Khattri, N., Reeve, A. L., Kane, M. B., and Adamson, R. J. (1996). *Studies of educational reform: Assessment of student performance.* Final report, volume I: Findings and conclusions. Washington, D.C.: Pelavin Research Institute

132

Knapp, J. E. (1995). *Building an integrated and quality skill standards certification system.* Washington, D.C.: Institute for Educational Leadership.

Linnehan, F. (1996). Measuring the effectiveness of a career academy program from an employer's perspective. *Educational Evaluation and Policy Analysis, 18*(1), 73-89.

Linn, Robert L. (1994). Performance assessment: Policy promises and technical measurement standards. *Educational Researcher,* 23(9), 4-12.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement.* Third Edition. New York, NY: Macmillan Publishing Company.

MPR Associates. (1996). *Skill standards: Concepts and practices in state and local education.* Berkeley, CA: MPR Associates, Inc.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Thousand Oaks, CA: SAGE Publications.

Nash, B. E. (1995). *The national job analysis study: A project to identify cross-occupational skills related to high-performance workplaces.* Iowa City, Iowa: ACT.

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, D.C.: U.S. Government Printing Office.

National Retail Federation. (1996). *Putting the standards to work: Performance indicators* (retail skill standards). Washington, D.C.: National Retail Federation.

National Skill Standards Board. (1996). *Voluntary skill standards: For employers, employees, educators, applicants and students.* Washington, D.C.: National Skill Standards Board.

National Skill Standards Board. (1996). *Proposal to establish a voluntary national skill standards system.* Washington, D.C.: National Skill Standards Board.

Northwest Regional Educational Laboratory. (1996). *What is work-based learning and how do we measure it: A toolkit for decisionmakers.* Portland, OR: NWREL.

Oregon Department of Education (Jonnerg, 1997). *Oregon Standards.* Draft. Salem, OR: Oregon Department of Education

Phillips, S. E. (1993). *Legal implications of high-stakes assessment: What states should know*. Oak Brook, IL: North Central Regional Educational Lab.

Rabinowitz, S. N. (1997). A comprehensive performance-based system to address work readiness. In O'Neil, H. (1996). *Workforce readiness: Competencies and assessment*. Lawrence Erlbaum Associates.

Resnick, L. B. and Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford and M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitute, achievement and instruction*. Boston: Kluwer.

Schniedman, Lambert, and Wander. (1989). *Workbook for being a nursing assistant*. Fifth edition. New Jersey: Prentice-Hall.

Schwager, M.T. (1997). *Developing cross-industry assessments: Using cross-industry categories to focus assessment development*. San Francisco, CA: WestEd.

Secretary's Commission on Achieving Necessary Skills (SCANS). (June, 1992). *SCANS in the schools*. Washington, D.C.: U.S. Department of Labor.

Shavelson, R. J., Baxter, G. P., and Gao, X. (1993). Sampling variability of performance assessments. *Journal of Education Measurement, 30*, 215-232.

Smith, S. (1996). *Skill standards development and implementation: Evaluation guide*. U. S. Department of Education.

Stetcher, B. M., Rahn, M. L., Ruby, A., Alt, M. N., Robyn, A., with Ward, B. (1997). *Using alternative assessment in vocational education*. Berkeley, CA: National Center for Research in Vocational Education.

Stern, D., Raby, M., and Dayton, C. (1992). *Career academies: Partnerships for reconstructing American high schools*. San Francisco, CA: Jossey-Bass Publishers.

WestEd. (1998). *A guide to standards-based assessment systems: The C-TAP example*. Sacramento, CA, California Department of Education and Sacramento County Office of Education.

Wheeler, P. and Haertel, G. D. (1993). *Resource handbook on performance assessment and measurement: A tool for students, practitioners, and policymakers*. Berkeley, CA: The Owl Press.

Wiggins, G. P. (1993). *Assessing student performance*. San Francisco, CA: Jossey-Bass Publishers.

Wills, J. (1997). *Standards: Making them useful and workable for the education enterprise*. Washington, D.C.: Office of Vocational and Adult Education, U.S. Department of Education.

Wirt, John G. (1993). *Systems of performance assessment: Implications for national system of industry skill standards*, Volume 1. Draft paper.

Woolfolk, A. E. (1993). *Educational psychology: Fifth edition*. Boston, MA: Allyn and Bacon.

NSSAC Project Staff

Srijati Ananda, PhD - Project Director

Mahna Schwager, PhD - Senior Research Associate

Michael Crepeau - Research Assistant

Burr Tyler - Research Associate

Justine McCarthy - Consultant

Bonnie Lurie - Administrative Assistant

Fredrika Baer - Administrative Assistant, desktop publishing